

**PATENT ABSTRACTS OF JAPAN**

(11)Publication number : 2000-076473

(43)Date of publication of application : 14.03.2000

---

(51)Int.Cl. G06T 11/60  
G06F 3/00  
G06F 12/00  
G06F 13/00

---

(21)Application number : 11-100108

(71)Applicant : FUJI XEROX CO LTD

(22)Date of filing : 07.04.1999

(72)Inventor : BICKMORE TIMOTHY W  
SCHILIT WILLIAM N  
GIRGENSOHN ANDREAS  
SULLIVAN JOSEPH W

(30)Priority

Priority number : 98 80909\_ Priority date : 07.04.1998 ; Priority country : US  
99 239295 29.01.1999 US

---

(54) AUTOMATIC DOCUMENT RE-AUTHORING METHOD AND SYSTEM THEREFOR

(57)Abstract:

PROBLEM TO BE SOLVED: To provide a document re-authoring method and system for displaying a web document which is designed to be displayed in a desktop computer on a small display of a PDA, a portable telephone or the like.

SOLUTION: In this document re-authoring system, the web document obtained from a distributed network is analyzed, an abstract syntax tree is generated, various kinds of deformation are applied to the element of the document, and the document is divided into small sub-pages capable of being deciphered and navigated on the device limited in display area. Besides, the system is provided with a document filtering sub-system to select the information to be displayed in the device limited in the display area by a user.

# JP 2000-076473 A

(11)Publication number : 2000-076473 (51)Int.Cl. G06T 11/60  
(43)Date of publication of application : 14.03.2000  
(21)Application number : 11-100108 (71)Applicant : FUJI XEROX CO LTD  
(22)Date of filing : 07.04.1999 (72)Inventor : BICKMORE TIMOTHY W  
SCHILIT WILLIAM N  
GIRGENSOHN ANDREAS  
SULLIVAN JOSEPH W

(30)Priority

Priority number : 98 80909 Priority date : 07.04.1998 Priority country : US

(54) AUTOMATIC DOCUMENT RE-AUTHORING METHOD AND SYSTEM THEREFOR

(57)Abstract:

PROBLEM TO BE SOLVED: To provide a document re-authoring method and system for displaying a web document which is designed to be displayed in a desktop computer on a small display of a PDA, a portable telephone or the like.

SOLUTION: In this document re-authoring system, the web document obtained from a distributed network is analyzed, an abstract syntax tree is generated, various kinds of deformation are applied to the element of the document, and the document is divided into small sub-pages capable of being deciphered and navigated on the device limited in display area. Besides, the system is provided with a document filtering sub-system to select the information to be displayed in the device limited in the display area by a user.

## Disclaimer

This is a machine translation performed by INPIT (<http://www.ipdl.inpit.go.jp>) and received and compiled with PatBot (<http://www.patbot.de>). PatBot can't make any guarantees that this translation is received and displayed completely!

## Notices from INPIT

Copyright (C) JPO, INPIT

The JPO and INPIT are not responsible for any damages caused by the use of this translation.

- 1.This document has been translated by computer. So the translation may not reflect the original precisely.
- 2.\*\*\*\* shows the word which can not be translated.
- 3.In the drawings, any words are not translated.

---

CLAIMS

---

[Claim(s)]

[Claim 1]Are the method of carrying out re-authoring of the document automatically, analyze a document, and an analyzed document is transformed into a changed document, When an evaluation value is generated from a changed document, it determines whether an evaluation value satisfies at least one valuation basis and an evaluation value to a changed document does not satisfy at least one valuation basis, An automatic document re-authoring method which outputs a changed document when modification, generation, and a determination step are repeated using different modification and an evaluation value to a changed document satisfies at least one valuation basis.

[Claim 2]Determine whether modification of an analyzed document can apply modification to a document in which it chose and selected modification was analyzed properly, and selected modification properly when it can apply, A method according to claim 1 of changing into a changed document using modification which had an analyzed document chosen, and repeating a step of selection and determination to different modification, when selected modification cannot apply properly.

[Claim 3]When there is no modification which produces a changed document which has an evaluation value with which it is satisfied of at least one valuation basis, A method according to claim 1 of choosing a changed document which has an evaluation value nearest to satisfying an evaluation value, and repeating modification, generation, and a determination step using additional modification in a selected changed document.

[Claim 4]A way according to claim 1 modification to a changed document of a document is at least one of the abstracts of outline-izing of a section of a document, removal of a portion without contents from a document, removal of contents from a document, reduction of at least one picture in a document, and a text in a document.

[Claim 5]Outline-ization of a section of a document identifies a section in a document, In order to form a changed document which identifies a section header and a document portion about each section, arranges each identified document portion to separate sub pages, and contains only an identified section header, A method according to claim 4 of removing a document portion discriminated from an analyzed document, changing each identified section header into a link to corresponding sub pages, and linking separate sub pages to mutual and a changed document.

[Claim 6]In order that reduction of at least one picture in a document may identify at least one picture in a document, may arrange each identified picture to separate sub pages, may create a reduction version of each identified picture and may form a changed document, A method according to claim 4 of inserting a reduction version of each picture removed by removing each identified picture from a document, and adding a link to sub pages which include the removed picture in a reduction version of the picture about each removed picture.

[Claim 7]A way according to claim 4 reduction of at least one picture in a document reduces further size of a picture reduced before.

[Claim 8]Removal of contents from a document is the method of removal of at least one picture from a document, and removal of a cell of at least one table from a document according to claim 4 which is either at least.

[Claim 9]A way according to claim 8 removal of at least one picture from a document is either of removal of all the pictures from a document, removal of all the pictures other than a picture of the beginning from a document, and removal of all the pictures other than a picture of the beginning from a document, and the last.

[Claim 10]A method according to claim 9 of replacing each identified picture with a link to corresponding sub pages, in order that removal of all the pictures from

a document may identify each picture in a document, may add each identified picture to separate sub pages and may form a changed document.

[Claim 11] In order that removal of all the pictures from a document may identify each picture in a document, may add each identified picture to separate sub pages and may form a changed document, A method according to claim 9 of replacing the 1st identified picture with a link to corresponding sub pages, removing a picture from which others were discriminated from a changed page, and linking separate sub pages mutually.

[Claim 12] A method according to claim 9 of replacing each identified picture with a link to corresponding sub pages, in order that removal of all the pictures other than a picture of the beginning from a document may identify each picture other than a picture of the beginning in a document, may add each identified picture to separate sub pages and may form a changed document.

[Claim 13] In order that removal of all the pictures other than a picture of the beginning from a document may identify each picture other than a picture of the beginning in a document, may add each identified picture to separate sub pages and may form a changed document, A method according to claim 9 of removing a picture which added one link in separate sub pages to the first picture, and was discriminated from a changed page, and linking separate sub pages mutually.

[Claim 14] A method according to claim 9 of replacing each identified picture with a link to corresponding sub pages, in order that removal of all the pictures other than a picture of the beginning from a document and the last may identify each picture other than a picture of the beginning in a document, and the last, may add each identified picture to separate sub pages and may form a changed document.

[Claim 15] In order that removal of all the pictures other than a picture of the beginning from a document and the last may identify each picture other than a picture of the beginning in a document, and the last, may add each identified picture to separate sub pages and may form a changed document, A method according to claim 9 of adding the 1st one link in separate sub pages to the first picture, and adding the 2nd one link in separate sub pages to the last picture, removing a picture discriminated from a changed page, and linking separate sub pages mutually.

[Claim 16] It is determined whether removal of a cell of at least one table from a document contains a sidebar of links with an arbitrary table, When a table contains arbitrary sidebars, in order to change a sidebar into a list of links as a cell of the last of a table, to identify all the cells other than a cell of the beginning of a table, to add each identified cell to separate sub pages and to form a changed document, A method according to claim 8 of replacing a table with the first cell and linking separate sub pages to mutual and a changed document.

[Claim 17] When it determines whether removal of a cell of at least one table from a document contains a sidebar of links with an arbitrary table and a table contains arbitrary sidebars, In order to change a sidebar into a list of links as a cell of the last of a table, to identify each cell of a table, to add each identified cell to separate sub pages and to form a changed document, A method according to claim 8 of replacing a table with one link in separate sub pages, and linking separate sub pages mutually.

[Claim 18] A way according to claim 1 modification to a changed document of an analyzed document creates at least one sub pages further.

[Claim 19] When a changed document satisfies at least one valuation basis, an evaluation value to each sub pages created to the changed document is generated further, It is determined whether, about each sub pages, an evaluation value to the sub pages satisfies at least one valuation basis, When an evaluation value to the sub pages does not satisfy at least one valuation basis about each sub pages, It changes into the sub pages using one of the modification additional in order to create changed sub pages, A method according to claim 18 of recognizing the sub pages as an output preparation completion, when generation and a

determination step are performed and an evaluation value to said sub pages satisfies at least one valuation basis about each sub pages.

[Claim 20]When changed sub pages satisfy at least one valuation basis, further, An evaluation value to each sub pages created to the changed sub pages is generated, It is determined whether, about each sub pages, an evaluation value to the sub pages satisfies at least one valuation basis, When an evaluation value to the sub pages does not satisfy at least one valuation basis about each sub pages, It changes into the sub pages using one of the modification additional in order to create changed sub pages, A method according to claim 18 of recognizing the sub pages as an output preparation completion, when generation and a determination step are performed and an evaluation value to the sub pages satisfies at least one valuation basis about each sub pages.

[Claim 21]When an evaluation value is generated from a document, it determines whether an evaluation value satisfies at least one valuation basis and a document does not satisfy at least one valuation basis after analysis of a document, A method according to claim 1 of outputting a document, without transforming a document, when modification, generation, and a determination step are performed using the first one of the modification and a document satisfies at least one valuation basis.

[Claim 22]In order to form a document changed [ reserve ] after analysis of a document, a portion without the contents is removed from a document, Generate an evaluation value from a document changed [ reserve ], and it is determined whether an evaluation value satisfies at least one valuation basis, When an evaluation value to a document changed [ reserve ] does not satisfy at least one valuation basis, A method according to claim 1 of outputting a document changed [ reserve ], without removing no contents from a document, when modification, generation, and a determination step are performed using the first one of the modification and an evaluation value to a document changed [ reserve ] satisfies at least one valuation basis.

[Claim 23]A method according to claim 1 of replacing with a portion from which modification of a document filtered a document in order to extract a portion of a request of a document, and a document was extracted.

[Claim 24]A document re authoring system which carries out re-authoring of the document automatically, comprising:

A parsing tree generating circuit.

A document size weighting network.

A modification circuit.

[Claim 25]The document re authoring system according to claim 24 which analyzes a document in order that a parsing tree generating circuit may generate a parsing tree.

[Claim 26]The document re authoring system according to claim 25 whose parsing tree is an abstract syntax tree.

[Claim 27]The document re authoring system according to claim 25 by which a parsing tree in which a document size weighting network was generated by parsing tree generating circuit is evaluated in order to determine whether a document satisfies at least one valuation basis.

[Claim 28]The document re authoring system according to claim 27 with which a document is outputted to a display which has a viewing area smaller than a viewing area of a desktop monitor when a document satisfies at least one valuation basis.

[Claim 29]The document re authoring system according to claim 27 which a modification circuit uses modification of the 1st, transforms a parsing tree, and generates the 1st modification parsing tree when a document does not satisfy at least one valuation basis.

## [Detailed Description of the Invention]

[0001]

[Field of the Invention] This invention carries out re-authoring of the arbitrary documents from World Wide Web automatically, in order to display suitable for a Personal Digital Assistant (PDA) and a small screen device like a cellular phone. It is related with the document re authoring system and method of providing access to the web which is not influenced by the device.

[0002]

[Description of the Prior Art] Access to the World-Wide-Web document from a personal electronic device, contest the Institute of Electrical and Electronic Engineers comp (IEEE COMPCON) -- 95 and San Francisco, California,. "The experience by a wireless World-Wide-Web client (Experience with a Wireless World Wide Web Client)" of J. Bartlett (J. Bartlett) in March, 1995, The 2nd international World-Wide-Web meeting (International World Wide Web Conference), Chicago, Illinois, "PDA as a mho BAIRU WWW browser (PDA as Mobile WWW Browsers)" of S. Gaessler and others (S. Gessler et al.) in October, 1994, The workshop (Workshop on Mobile Computing Systems and Applications) about a MOBA yl computing system and application, California Santa Cruz, G. Volcker et al. in December, 1994 (G.) "mho by ZAIKU of Voelker et al.: The information system (Mobisaic:An Information System for a Mobile.) for MOBA yl wireless computing environment Wireless Computing Environment", And a 1994 MOBA yl computing system and an application workshop (Mobile Computing Systems and Applications Workshop) position paper, T. Watson in August, 1994. (T. It has demonstrated by a research project which is indicated in "the application design (Application Design for Wireless Computing) for wireless computing" of Watson). Such access is reality commercial now. Presto of the general magic (General Magic) corresponding to Magic Link (MagicLink) of Sony! Links (Presto!Links), Each net hopper (NetHopper) of all pen (AllPen) corresponding to MI-10 of Newton and sharp (Newton and Sharp) provides the WWW browser for a PDA class device.

On the other hand, the duet (Duett) of Nokia 9000 communicator (Nokia9000 Communicator) and Samsung (Samsung) provides the capability to access a web from a cellular phone.

[0003] Though regrettable, almost all the pages on World Wide Web and other distributed networks are designed so that resolution may display on the desktop computer which has a color monitor of at least 640x480. Many pages may be designed supposing the monitor of bigger resolution. On the other hand, the device of almost all the PDA class and the display of a cellular phone are far small. If the ratio of the screen field of the designed screen field versus a hand obtains from 4 to 1 to 100 to 1 (or more greatly than it) and displays a World-Wide-Web document directly on these small devices according to the difference in this viewing area, Esthetically, undesirably, it cannot navigate, and most completely becomes decipherment impossible, when the worst. The central problem in access to the worldwide web page for which this used these small devices, That is, the problem how to display on the personal electronic device which has the display ability to which arbitrary documents like the HTML document designed for desktop systems were restricted far is presented.

[0004] Although art has already provided the mobility and wireless connectivity of a computer, The standard solution for seeing a document and a web page on a small screen, When the user is carrying the magnifying glass by chance, it is equipping with a facsimile or printing capacity the increase in the screen resolution which can be referred to as wonderful, or familiar hard copy equipment, but both are inconvenient and it is contradictory to the rationality of having an electronic filing document in the first place. There are the five general methods of displaying a web document on a small screen device. That is, they are device specification authoring, authoring corresponding to a multi-device, client side navigation, automatic re-authoring, and web page filtering. Device specification authoring includes authoring of 1 set of special web documents for displays like

a cellular phone which equipped a display and communicating software, such as Nokia 9000, for example. The user of such a special device only intends to access the set from which service was chosen, and the fundamental view of this approach has him. Therefore, the document for these services should be designed by the tip corresponding to the special display system of the device to access. Although information may be provided from a general distributed network, the desired page must be extraction of fixed information and custom information, and page-formats software should be written in order to display information on a small device. This is the method which UP link (UP.Link) service of the ANWAIADO planet (Unwired Planet) using the markup language (HDML) of monopoly has taken.

[0005] In authoring corresponding to a multi-device, the range of the device made into the purpose is specified. And since the device within the specified limits is covered, mapping to 1 set of display documents from a single source document is defined. This one example The Kent university of a Canterbury computing laboratory WWW page (University of Kent at Canterbury Computing Laboratory WWW Page), I. Cooper et al. in November, 1995 (I.) "PDA browser of Cooper et al.: It is the stretch text (StretchText) approach currently discussed by realization problem (PDA Web Broesers: Implementation Issues)." In a stretch text, even a word level goes down to the partial portion of a document potentially, and the level of abstract measure can be added to it. When receiving a document, the user can specify the abstract level which he wants to see, and a document is displayed with the detailed or detailed lack corresponding to specification.

[0006] Another example of authoring corresponding to a multi-device, It is a HTML cascading style sheet (CSS) like the statement to "a cascading style sheet (Cascading Style Sheets)" of WWW associations (WWW Consortium) and H. rye (H. Lie et al.) in September, 1996. By a cascading style sheet, a single style sheet specifies 1 set of display attributes corresponding to the structured division from which a document differs. For example, it can be specified that all the section headers of the highest level are displayed in red a TAIMUZU (Times) font and 18 point. A series of style sheets which have the dignity (weight) each style sheet describes the desirability of the style sheet for a document preparation person to be may be attached to one document. The user can also specify a default style sheet. In order to access a distributed network, the browser used by the user can also specify a "default" style sheet. Usually, although a maker's style sheet disregards a user's style sheet, the user can make a maker's style sheet effectively or invalid selectively, and can equip a user's specific display with the capability to which drawing of a document is fitted.

[0007] In client side navigation, a user is changing the portion currently displayed on the arbitrary time of the single web page, and can give the capability to navigate the inside of a single web page to a dialogue. This serious trifling example is use of the scroll bar of a document display field. Association for Computing Machinery (ACM) UIST'94 conference note, an ACM press (ACM Press), B. BIDASON et al. in 1994 (B.) "Pad++ of Bederson et al.: The graphical interface (Pad++:A Zooming Graphical Interface.) for investigation of alternative interface physics which carries out zoom In a PAD(packet assembly/decomposition)++ system like a statement, to for ExploringAlternate Interface Physics." Zoom and an approach which can carry out pan (move a screen to right and left) and which was refined far are freely taken in the display of the device from one end of the document in which a user has infinite resolution to the other. The 2nd international World-Wide-Web meeting, Chicago, Illinois, J. Huu et al. in October, 1994 (J.) Formation of active outline for "HTML document of Hsu et al. : like a statement to realization (Active Outlining for HTML Documents: An X-Mosaic Implementation)" of X-mosaic, Active outline-ization is also carried out as client side navigation technology, and a user can expand and minimize the section of a document dynamically under an individual section header in this art (collapse). Otherwise, the art included in this category Computer human interaction (Computer-Human Interaction):CHI96, a conference note, Canada,

the State of British Columbia, Vancouver, A translucent mechanism like the statement to "the more efficient use (Using small screen space more efficiently) of small screen space" of T. Kamba and others (T. Kamba et al.) in April, 1996, "tool glass and a magic lens of E. Bahia and others (E. Bier et al.) in SIGGRAPH'93 conference-note 1993: A see-through interface (Toolglass and Magic Lenses: .) A magic lens system like a statement is included in The See-through Interface."

[0008]Automatic document re-authoring arbitrary documents like the HTML document designed so that it might be displayed on the monitor of desktop size, The developing software which can carry out re-authoring of the arbitrary documents via a series of modification is included so that it may incorporate with the characteristic of a target display and arbitrary documents can be appropriately displayed on a target display. This process can be performed by all of a relay proxy server like the HTTP proxy server which exists only for a client, a server, or the purpose of providing this modification service. The example of the approach of this latter The 5th international World-Wide-Web meeting, France, Paris, A. Fox et al. in May, 1996 (A.) Fox et al. "reduction (Reducing WWW Latency and Bandwidth Requirements.) of the World-Wide-Web waiting time by real-time distillation, and band requests By the Pythia (Pythia) proxy server of University of California at Berkeley of a statement, this proxy server transforms a web page picture into by Real-Time Distillation." However, the Pythia proxy server is hit only to minimization of page reading time in the focus. Spy glass prism (Spyglass Prism) is goods which perform automatic re-authoring of an HTML document using the fixed modification relevant to page tags or the embedded object type. For example, prism will reduce all the JPEG images 50%.

[0009]Finally, web page filtering enables a user to see only a portion with the interest of a page. In order to save a non-line zone region and an equipment memory, filtering may be performed on a relay server like a HTTP proxy server. However, filtering can be performed also with a client apparatus as display management engineering. The specification of a filter can be based on a keyword, regular representation collation or page structure navigation, and a global command. Filtering can be specified even if it uses any of a visual tool or a description language.

[0010]

[Problem(s) to be Solved by the Invention]In five approaches, device specification authoring, authoring corresponding to a multi-device, client side navigation, automatic re-authoring, and web page filtering, each has the strong point and demerit. Since human being's designer is engaged directly, device specification authoring brings about a result with the typical most sufficient appearance. However, device specification authoring limits a user's access to the small-scale selected set of the document by which authoring was carried out for the specific device. Although the total labor as which authoring corresponding to a multi-device is required per document is smaller than device specification authoring, Quite many design works by handicraft are needed rather than still carrying out authoring of the single version of the single document for desktop type plat forms simply. Good success will be achieved when the client side navigation can develop the good set of view (viewing) art. However, the client side navigation needs to distribute the whole document to a client apparatus simultaneously, and will waste a precious non-line zone region and memory. "inspection hole "approach (peephole) taken by PAD++, So that it may be very inconvenient to use it for a big document, and since a section / subsection composition with almost all strict web pages are not used for active outline-ized art or is incorrectly used for it, applicability is restricted.

[0011]Therefore, if automatic re-authoring which draws up a re-authoring document possible [ the decipherment possibility of and navigation ] and esthetically desirable can be made without losing information, automatic re-authoring, It is the ideal approach which provides large access to a web document and other web



contents from a wide range device.

[0012]This invention provides the system and method of carrying out re-authoring automatically, in order to display the document designed according to the bigger viewing area on a smaller viewing area.

[0013]This invention provides the system and method which a viewing area required for each sub documents transforms into two or more smaller sub documents to which it was linked for a document automatically.

[0014]This invention provides the system and method of applying several different modification to the original document automatically, in order to create two or more sets of the linked sub documents.

[0015]This invention provides further the system and method of applying several different modification to at least one of two or more of the sets of the linked sub documents automatically, in order to draw up the sub documents in which the addition was linked.

[0016]This invention provides further the system and method of analyzing the main sub documents of each set of the linked sub documents, in order to determine one of best [ the / of main sub documents ].

[0017]If it cannot opt for and indicate whether the further best main sub documents can display this invention on a smaller viewing area, in order to reduce a required viewing area further, the system and method of applying modification to the main sub documents further are provided.

[0018]This invention provides the system and method of filtering a document, in order to extract the portion of a request of the document which can be displayed on a smaller viewing area.

[0019]This invention provides the system and method of filtering a document, in order to extract the portion described based on the predetermined script.

[0020]This invention provides document filtering for extracting a desired portion with the system and method of creating an usable script.

[0021]This invention provides writing the script for document filtering for extracting a desired portion with an usable script language.

[0022]

[Means for Solving the Problem]In one illustration embodiment, a document re authoring system and a method of this invention, In order to obtain a document with the most sufficient appearance compared with given display size, it realizes on a HTTP proxy which uses heuristics planning art and 1 set of structural page modification, and carries out re-authoring of the demanded web page dynamically. Automatic document re-authoring according to a system and a method of this invention can be performed by client, server, or one illustration embodiment by all of a relay HTTP proxy server which exists only for the purpose of providing these modification services. An automatic document re authoring system and a method of having followed this invention can be performed even if it combines these devices.

[0023]An automatic document re authoring system and a method of this invention achieve success as which a display looked at by PDA may be sufficient. However, when a document re authoring system and a method of this invention are applied to a very much limited display looked at by the present cellular phone, a document re authoring system and a method of this invention create a page with sometimes difficult navigation. When accessing a distributed network like the Internet or intranet from a cellular phone, almost all users are interested in access to information mainly specified very much. A document filtering system and a method of this invention provide those users with manual control which limits information which you want to display. A document filtering system and a method of this invention return easily only a small portion of a page which can be navigated. Since a filter is set by format of a page, a user is in a situation where a user is monitoring a specific page where a layout is fixed and which changes the contents, and a document filtering system and a method of this invention have him. [ ideal ]

[0024]Automatic document re-authoring, a document filtering system, and a method

of this invention, In order to provide with access to arbitrary documents on a distributed network like the Internet or intranet a device which has a limited communication band and a small display, automatic document re-authoring capability connected with document filtering is provided.

[0025]Automatic document re-authoring, a document filtering system, and a method of this invention carry out interception (intercept) of the demand to a document from a distributed network, and return a re-authoring finishing version of a demanded document instead of a document of demanded origin.

[0026]In a bigger context [ say / mho BAIRU and ubiquitous (ubiquitous) computing ]. Automatic document re-authoring, a document filtering system, and a method of this invention provide a user with important art of giving view mobility (view-mobility) crossed to various plat forms.

[0027]The concrete mode of this invention is as follows.

[0028]The 1st mode of this invention is the method of carrying out re-authoring of the document automatically, Analyze a document, transform an analyzed document into a changed document, and an evaluation value is generated from a changed document, When it determines whether an evaluation value satisfies at least one valuation basis and an evaluation value to a changed document does not satisfy at least one valuation basis, When modification, generation, and a determination step are repeated using different modification and an evaluation value to a changed document satisfies at least one valuation basis, it is the automatic document re-authoring method which outputs a changed document. In the 1st mode, as for the 2nd mode, an output of a changed document transmits a changed document to a display. The 3rd mode has a viewing area in which a display is smaller than a viewing area of a desktop monitor in the 2nd mode. In the 1st mode, as for the 4th mode, analysis of a document generates an abstract syntax tree from a document. In the 4th mode, modification of an analyzed document transforms the 5th mode into at least one modification abstract syntax tree for an abstract syntax tree. Modification of a document in which the 6th mode was analyzed in the 1st mode chooses modification, Determine whether it is properly applicable to a document in which selected modification was analyzed, and selected modification properly when it can apply, It changes into a changed document using modification which had an analyzed document chosen, and when selected modification cannot apply properly, a step of selection and determination is repeated to different modification. Modification as which determination of whether the 7th mode is properly applicable to a document in which selected modification was analyzed in the 6th mode was chosen is the determination of whether to be contradictory to modification applied before. The 8th mode is the determination of whether determination of whether to be properly applicable to a document in which selected modification was analyzed satisfies an application standard over modification as which an analyzed document was chosen in the 6th mode. It is at least one of the abstracts of outline-izing of a section of a document, removal of contents from a document, reduction of at least one picture in a document, and a text in a document that the 9th mode changes into a changed document using modification which had an analyzed document chosen in the 6th mode. When the 10th mode does not have modification which produces a changed document which has an evaluation value with which it is satisfied of at least one valuation basis in the 1st mode, A changed document which has an evaluation value nearest to satisfying an evaluation value is chosen, and modification, generation, and a determination step are repeated using additional modification in a selected changed document. In the 1st mode, modification to a changed document of a document of the 11th mode is at least one of the abstracts of outline-izing of a section of a document, removal without contents from a document of a portion, removal of contents from a document, reduction of at least one picture in a document, and a text in a document. In the 11th mode, the 12th mode outline-ization of a section of a document, Identify a section in a document, arrange each document portion which identified a section header and a document portion about each section, and was identified to separate sub pages, and in order to form a changed document containing only an identified section header, A

document portion discriminated from an analyzed document is removed, each identified section header is changed into a link to corresponding sub pages, and separate sub pages are linked to mutual and a changed document. In the 12th mode, discernment of a section of the 13th mode is discernment of a text block in a document. In the 13th mode, discernment of a section header in a text block and a document portion makes a typical text string of a text block a section header, and the 14th mode chooses a text block as a document portion. In the 14th mode, a text string of the 15th mode is at least a part of 1st sentence of a text block. In the 14th mode, a text string of the 16th mode is a section header of a text block. Removal of a portion in which the 17th mode does not have the contents from a document in the 11th mode is replacing comparatively by single page division or a paragraph about a sequence of page division or paragraph division. The 18th mode is removal of a format of removal of a portion without contents from a document from a document in the 11th mode. The 19th mode changes removal of an indent of removal of a format from a document from a document, and a text string of a document into at least one of a single font and the single sizes in the 18th mode. It is at least one of removal of a black dot from a document, removal of a background space from a document, and removal of banner images from a document. The 20th mode replaces with a text link where removal of banner images from a document corresponds banner images further in the 19th mode. In the 11th mode, the 21st mode reduction of at least one picture in a document, In order to identify at least one picture in a document, to arrange each identified picture to separate sub pages, to create a reduction version of each identified picture and to form a changed document, A reduction version of each picture removed by removing each identified picture from a document is inserted, and a link to sub pages which include the removed picture in a reduction version of the picture is added about each removed picture. The 22nd mode reduces size of a picture by which reduction of at least one picture in a document was reduced further before in the 11th mode. In order that reduction of size of a picture reduced before may identify a picture reduced before [ in a document ] at least one and may form a changed document in the 22nd mode, the 23rd mode, Each picture reduced before being discriminated from a document is changed into a version to which each picture reduced before was reduced further. In the 11th mode, removal of contents from a document of the 24th mode is either at least as removal of at least one picture from a document, and removal of a cell of at least one table from a document. In the 24th mode, removal of at least one picture from a document of the 25th mode is either of removal of all the pictures from a document, removal of all the pictures other than a picture of the beginning from a document, and removal of all the pictures other than a picture of the beginning from a document, and the last. In the 25th mode, the 26th mode replaces each identified picture with a link to corresponding sub pages, in order that removal of all the pictures from a document may identify each picture in a document, may add each identified picture to separate sub pages and may form a changed document. The 27th mode links separate sub pages about each identified picture mutually further in the 26th mode. The 28th mode contains either of a text string relevant to each picture from which each link was discriminated, and a predetermined icon showing a picture in the 26th mode. The 29th mode comes to hand in the 28th mode from hypertext information relevant to each picture from which a text string relevant to each identified picture was discriminated. In order that removal of all the pictures from a document may identify each picture in a document, may add each identified picture to separate sub pages and may form a changed document in the 25th mode, the 30th mode, The 1st identified picture is replaced with a link to corresponding sub pages, a picture from which others were discriminated is removed from a changed page, and separate sub pages are linked mutually. The 31st mode contains either of a text string relevant to each picture from which a link was discriminated, and a predetermined icon showing a picture in the 30th mode. The 32nd mode comes to hand in the 31st mode from hypertext information relevant to each picture from which a text string relevant to each identified picture was discriminated. In order that removal of all the pictures

other than a picture of the beginning from a document may identify each picture other than a picture of the beginning in a document, may add each identified picture to separate sub pages and may form a changed document in the 25th mode, the 33rd mode, Each identified picture is replaced with a link to corresponding sub pages. The 34th mode links separate sub pages about each identified picture mutually further in the 33rd mode. The 35th mode contains either of a text string relevant to each picture from which each link was discriminated, and a predetermined icon showing a picture in the 33rd mode. The 36th mode comes to hand in the 35th mode from hypertext information relevant to each picture from which a text string relevant to each identified picture was discriminated. In order that removal of all the pictures other than a picture of the beginning from a document may identify each picture other than a picture of the beginning in a document, may add each identified picture to separate sub pages and may form a changed document in the 25th mode, the 37th mode, One link in separate sub pages is added to the first picture, a picture discriminated from a changed page is removed, and separate sub pages are linked mutually. In the 25th mode, the 38th mode removal of all the pictures other than a picture of the beginning from a document and the last, In order to identify each picture other than a picture of the beginning in a document, and the last, to add each identified picture to separate sub pages and to form a changed document, each identified picture is replaced with a link to corresponding sub pages. The 39th mode links separate sub pages about each identified picture mutually further in the 38th mode. The 40th mode contains either of a text string relevant to each picture from which each link was discriminated, and a predetermined icon showing a picture in the 38th mode. The 41st mode comes to hand in the 40th mode from hypertext information relevant to each picture from which a text string relevant to each identified picture was discriminated. In the 25th mode, the 42nd mode removal of all the pictures other than a picture of the beginning from a document and the last, In order to identify each picture other than a picture of the beginning in a document, and the last, to add each identified picture to separate sub pages and to form a changed document, The 1st one link in separate sub pages is added to the first picture, and the 2nd one link in separate sub pages is added to the last picture, a picture discriminated from a changed page is removed, and separate sub pages are linked mutually. In the 24th mode, the 43rd mode removal of a cell of at least one table from a document, When it determines whether a table contains a sidebar of arbitrary links and a table contains arbitrary sidebars, In order to change a sidebar into a list of links as a cell of the last of a table, to identify all the cells other than a cell of the beginning of a table, to add each identified cell to separate sub pages and to form a changed document, a table is replaced with the first cell and separate sub pages are linked to mutual and a changed document. In the 24th mode, an addition to sub pages with each separate cell the 44th mode about each cell, It determines whether the cell is the table made into a nest, when the cell is not the table made into a nest, said cell is added to separate sub pages, and when the cell is the table made into a nest, determination of the 43rd mode, conversion, discernment, an addition, replacement, and links tetraethylpyrophosphate are repeated. In the 24th mode, the 45th mode removal of a cell of at least one table from a document, When it determines whether a table contains a sidebar of arbitrary links and a table contains arbitrary sidebars, In order to change a sidebar into a list of links as a cell of the last of a table, to identify each cell of a table, to add each identified cell to separate sub pages and to form a changed document, a table is replaced with one link in separate sub pages, and separate sub pages are linked mutually. In the 43rd mode, an addition to sub pages with each separate cell the 46th mode about each cell, It determines whether the cell is the table made into a nest, when the cell is not the table made into a nest, said cell is added to separate sub pages, and when the cell is the table made into a nest, determination of the 43rd mode, conversion, discernment, an addition, replacement, and links tetraethylpyrophosphate are repeated. Modification to a changed document of a document in which the 47th mode

was analyzed in the 1st mode creates at least one sub pages further. In the 47th mode, when a changed document satisfies at least one valuation basis, the 48th mode, An evaluation value to each sub pages created to the changed document is generated, It is determined whether, about each sub pages, an evaluation value to the sub pages satisfies at least one valuation basis, When an evaluation value to the sub pages does not satisfy at least one valuation basis about each sub pages, It changes into the sub pages using one of the modification additional in order to create changed sub pages, When generation and a determination step are performed and an evaluation value to said sub pages satisfies at least one valuation basis about each sub pages, the sub pages are recognized as an output preparation completion. In the 48th mode, recognition as an output preparation completion of sub pages of the 49th mode is storing into output cash of the sub pages. In the 47th mode, when changed sub pages satisfy at least one valuation basis, the 50th mode, An evaluation value to each sub pages created to the changed sub pages is generated, It is determined whether, about each sub pages, an evaluation value to the sub pages satisfies at least one valuation basis, When an evaluation value to the sub pages does not satisfy at least one valuation basis about each sub pages, It changes into the sub pages using one of the modification additional in order to create changed sub pages, When generation and a determination step are performed and an evaluation value to the sub pages satisfies at least one valuation basis about each sub pages, the sub pages are recognized as an output preparation completion. The 51st mode generates an evaluation value from a document after analysis of a document further in the 1st mode, When it determines whether an evaluation value satisfies at least one valuation basis and a document does not satisfy at least one valuation basis, modification, generation, and a determination step are performed using the first one of the modification, and when a document satisfies at least one valuation basis, a document is outputted, without transforming a document. In the 1st mode, further, in order to form a document changed [ reserve ] after analysis of a document, the 52nd mode, Remove a portion without the contents from a document and an evaluation value is generated from a document changed [ reserve ], When it determines whether an evaluation value satisfies at least one valuation basis and an evaluation value to a document changed [ reserve ] does not satisfy at least one valuation basis, When modification, generation, and a determination step are performed using the first one of the modification and an evaluation value to a document changed [ reserve ] satisfies at least one valuation basis, a document changed [ reserve ] is outputted without removing no contents from a document. Removal of a portion in which the 53rd mode does not have the contents from a document in the 52nd mode is replacing comparatively by single page division or a paragraph about a sequence of page division or paragraph division. The 54th mode is removal of a format of removal of a portion without contents from a document from a document in the 52nd mode. The 55th mode changes removal of an indent of removal of a format from a document from a document, and a text string of a document into at least one of a single font and the single sizes in the 54th mode, It is at least one of removal of a black dot from a document, removal of a background space from a document, and removal of banner images from a document. The 56th mode replaces with a text link where removal of banner images from a document corresponds banner images further in the 55th mode. In the 1st mode, modification of a document filters a document, in order to extract a portion of a request of a document, and the 57th mode replaces with a portion from which a document was extracted.

[0029]The 58th mode of this invention is a document re authoring system which carries out re-authoring of the document automatically, and is a document re authoring system which has a parsing tree generating circuit, a document size weighting network, and a modification circuit. In the 58th mode, the 59th mode analyzes a document, in order that a parsing tree generating circuit may generate a parsing tree. In the 59th mode, a parsing tree of the 60th mode is an abstract syntax tree. In the 59th mode, the 61st mode evaluates a parsing tree in which a document size weighting network was generated by parsing tree generating circuit,

in order to determine whether a document satisfies at least one valuation basis. A document is outputted to a display in which the 62nd mode has a viewing area smaller than a viewing area of a desktop monitor when a document satisfies at least one valuation basis in the 61st mode. In the 61st mode, when a document does not satisfy at least one valuation basis, a modification circuit uses modification of the 1st, transforms the 63rd mode for a parsing tree, and it generates the 1st modification parsing tree. In the 63rd mode, the 64th mode evaluates a modification parsing tree in which a document size weighting network was generated by modification circuit, in order to determine whether a changed document corresponding to a modification parsing tree satisfies at least one valuation basis. In the 64th mode, when a changed document does not satisfy at least one valuation basis, a modification circuit uses modification of the 2nd, transforms the 65th mode for a parsing tree, and it generates the 2nd modification parsing tree. A changed document is outputted to a display in which the 66th mode has a viewing area smaller than a viewing area of a desktop monitor when a changed document satisfies at least one valuation basis in the 64th mode. the 67th mode answers modification of a parsing tree in the 63rd mode, and a modification circuit corresponds to at least one sub pages -- a parsing tree of sub pages is also generated at least. When a changed document satisfies at least one valuation basis in the 67th mode, in order to determine whether sub pages corresponding to each sub-pages parsing tree satisfy at least one valuation basis, the 68th mode, A document size weighting network evaluates a sub-pages parsing tree of each result generated by modification circuit from the changed document. When sub pages on the 68th mode and corresponding to [ tree / of each result / sub-pages parsing ] the sub-pages parsing tree in the 69th mode satisfy at least one valuation basis, the sub pages are recognized as an output preparation completion to a display. In the 68th mode, the 70th mode about a sub-pages parsing tree of each result. When sub pages corresponding to the sub-pages parsing tree do not satisfy at least one valuation basis, in order to generate a changed sub-pages parsing tree, a modification circuit uses modification of the 2nd and transforms the sub-pages parsing tree.

[0030]The feature and the strong point of the above of this invention and others are explained by detailed explanation of following desirable embodiments, or become clear from it.

[0031]

[Embodiment of the Invention]In the argument of the following document re-authoring of this invention, a document filtering system, and a method. The term of a "web page", a "web document", and a "document", It means including the set of arbitrary information searched as a simple substance from distributed networks, such as World-Wide-Web portions of intranet, the Internet, and the Internet, or other arbitrary distributed networks which are publicly known or were developed recently. This information may include a text string, a picture, a text string and the table of a picture, the link to another web page and the text string in a web page, a picture, a table, and the format information that specifies the layout of a link.

[0032]Many potential automatic document re-authoring technology exists, and they can be classified into syntactic opposite semantic art and modification pair abbreviation art in accordance with two dimensions. Syntactic art acts on the structure of a document and, on the other hand, depends for semantic art on a certain amount of contents understanding. Fundamentally, abbreviation art removes a certain information, and leaves except [ its ] as it is, and, on the other hand, modification art includes changing a certain mode of the method of presentation of a document, or the contents. Table 1 shows these dimensions with the example of each category.

[Table 1]

	省略	変形
構文的	セクションの アウトライン化	画像の縮小
意味的	重要でない内容の 除去	テキストの要約

異なるタイプの自動再オーサリング技術の例

[0033] In order to understand a process required for the automated document re authoring system, the characteristic of the typical web page was evaluated and the research which specifies the candidate of re-authoring technology was made via the process of carrying out re-authoring of some web pages by handicraft.

[0034] The Xerox [ which is a comparatively small-scale set of a web page "typical" in order to double the focus of research first ] corporate website (Xerox Corporate website) was chosen. A set of this web page of 3,188 is a sample of the site which the specialist of the latest style designed. The various statistics about these pages were collected using a web crawler (web crawler) as assistance for acquiring the structure of a typical page, and contents understanding. These statistics are in agreement with other large-scale researches done over the whole web in general.

[0035] Next, the subset of the page in the Xerox website was chosen as manual re-authoring. The Xerox [ in the 1995 fiscal year ] annual report (Xerox 1995 Annual Report) was chosen, and in order to display on Zaurus (Zaurus) PDA of sharp (Sharp) which has a screen which is 320x240 pixels, it was changed manually. The details of the used design strategy and art were recorded.

[0036] The followings are some of heuristics (heuristic programming) on the design mastered in this process.

- The thing of an original image for which it leaves some at least is important for maintaining sensibility of the appearance of a script. General art includes leaving, the picture, i.e., the bookends picture, of the first picture or the beginning, and the last, and omitting others.
  - The tags from H1 to H6 of a section header, i.e., HTML, are not used not much correctly. Even if the section header is used, there will be many people by whom it is used in order to obtain the specific font size and style like a board, for example. Therefore, not almost all documents can make a section header reliance providing structural outline. Instead, the document which has many text blocks is reducible by replacing each text block in the 1st sentence or the 1st phrase of each block, i.e., the 1st sentence abbreviation.
  - A picture is the percentage of the standard first decided by the ratio of a viewing area as authoring of the document was carried out, and the viewing area of a target device roughly, and reduces total image size. However, only a slight quantity can reduce a picture including a character or a number in the range whose decipherment of the contents does not become impossible.
  - A semantic abbreviation can be performed in the sidebar which displays the information from which it separated from the main concepts currently displayed into the page. Many of pages of Xerox have such a sidebar, and they were only omitted by the reduced version.
  - No information is given to a page but a semantic abbreviation can be carried out also to the picture of only the role which raises a fine sight.
  - A page can be classified into a category and can carry out re-authoring based on the category. These two examples are a banner and a table of a link. A banner is provided with the following.
- 1 set of pictures which almost or completely do not have the contents fundamentally only by the role which establishes a fine sight.  
It is often a navigation link of only one small number.  
When space is precious, this banner can usually be omitted thoroughly. The pages of the table of a link are 1 set of hypertext links linked to another page

fundamentally, therefore additional contents are hardly included. Usually, the page of the table of this link can be reformatted in the compacter form of only enumerating links in a text block.

- space natural on a big display is precious with a small device. Some art of reducing the quantity of the space in a page was discovered. The sequence of "P" "BR" tag of a paragraph, i.e., HTML, can be summarized to such one a paragraph or line feed. [ of a tag or line feed, i.e., HTKL, ] A list, i.e., "UL" of HTML and "OL", and/or the "DL" tag take a precious horizontal space by the indent or a black dot. As Cooper and others (Cooper et al.) has indicated, these lists can be reformatted to the simple text block which has line feed between continuous items.

[0037]In conclusion, in order to perform document re-authoring, two matters of the strategy which applies 1 set of re-authoring technology, i.e., 1 set of page modification, and page modification are required. What it is the easiest to systematize among the art in which it was used by manual re-authoring research is syntactic modification art including syntactic abbreviation art including the formation of section outline, the 1st sentence abbreviation, and a picture abbreviation, image size reduction, and font size reduction. the design strategy learned during research -- the rank of modification art -- namely, -- " -- the conditions which are 1 set by which this should be applied to the combination of trial" and each modification, or modification in front of that were included.

[0038]According to the research result discussed above, two main elements exist in the document re-authoring software system and method of this invention. That is, it is the re-authoring system and method of realizing a design strategy which were automated by choosing the technical best combination for a set of the individual re-authoring technology which transforms a document by various methods, and the pair of given document/display size.

[0039]Section header outline-ized modification (Section Header Outlining transform) provides reduction of the required display size of a document with a clear structure like engineering documentation and a report with a very good method. The outline-ized process is shown in drawing 1.

[0040]As shown in drawing 1, the document 100 is changed into the page 110 of a section list, and each section is omitted to the page 111. That is, the contents 106 of each section 102 of the document 100 are omitted from the document 100, and each section 104 is changed into a hypertext link. Selection of the hypertext link to arbitrary sections will load the page 111 of the omitted contents corresponding to it to a browser. When multiplex section levels (a section, a subsection, a subsubsection, etc.) are faced, two approaches exist in performing an abbreviation. The 1st approach works by leaving only a section header and omitting all the contents by all the outline-ization, and the result looks like the table of contents of a book. the 2nd approach responded to the level -- outline (to-level) -- it is-izing. In outline-ization according to a level, the cutoff level in a section hierarchy is determined, all the contents containing the section header of a lower level below the level are omitted, and all the contents above the level are left behind.

[0041]Since almost all pages have a text block, even when a section header does not exist, the 1st sentence abbreviation modification (First Sentence Elision transform) can serve as the good method of reducing a required screen field. In this art, each text block is replaced with the 1st phrase to a certain natural pause as that 1st sentence or substitution. This 1st sentence or phrase is also made into the hypertext link linked to an original text block.

[0042]Index segment modification (Indexed Segment transform) first tries discovery of a logically dividable page element as shown in alignment or a non-aligning list, the sequence of a paragraph, or a table. This modification incorporates the inputted page, divides the contents into sub pages by assigning some items to each sub pages, and builds and prepares the index page to a set of sub pages. Next, index segment modification begins to fill an output page with these elements in order until each page fills to the display size of a client.



When a single logical element is not settled in a single output page, index segment modification performs the secondary division which divides a text block based on the boundary of a paragraph or a sentence.

[0043] In index segment modification, as much style information over an output element as possible is held by outputting each element embedded in the HTML tag of all the ancestor divisions. And index segment modification copies a section header or the 1st sentence from the element each outputted. An index page is constituted by connecting the copied portion on an index page and creating the hypertext link from each copied portion to suitable sub pages. It should be recognized that the index page itself may be divided. In index segment modification, the "next" between continuous sub pages and a "pre-" navigation link are also added so that conveniently [ navigation ].

[0044] Front modification (Table transform) recognizes the case where the method of presentation of the information arranged by the table on a page, i.e., a rectangular grid, cannot be directly sent to a client. In this case, in the bottom, front modification creates one sub pages for every cell of a table using an order of the left to the right from a top. The table made into the nest into the table is also processed by the same method. Front modification judges the case where the sequence of the table currently performed by a commercial HTML web page being sufficient is used as a "navigation sidebar", using heuristics. In this case, since these cells tend to have almost no contents, front modification moves these cells to the end of the list of sub pages.

[0045] Drawing 2 makes the frame of a table thicker than the frame of the cell of a table, and shows the table made into the nest. In Table 120 shown in drawing 2, the cell 122 will be recognized as a sidebar and will be arranged after the cell 128. All other cells are arranged in a natural order. Six portions like the sub cells 125 and 126 of the cell 124 are arranged at the respectively original sub pages between the sub pages containing the sub cells 123 and 127, unless they contain only unfilled space.

[0046] The table and sidebar which were made into the nest complicate processing of a table so that an example may show. When a sidebar is a part of inside table, it is still more so. In the situation, the sidebar should be moved to the last of the inside table instead of the last of the table of the arbitrary circumferences. According to one illustration embodiment of the document re authoring system of this invention, and a method, without carrying out grouping of the cell by a table, a sidebar is moved about one table per time, and the cell of all the tables is processed at a time.

[0047] A picture presents one of the problems most difficult for automatic document re-authoring. It is because the determination of whether it leaves arbitrary pictures, it reduces, or to omit should be based on the contents of the picture on a page, and role understanding. However, as long as the mechanism in which a user can take out an original image is given, picture reduction modification (Image Reduction transform) and the picture abbreviation modification (Image Elision transform) cannot understand the contents, but \*\* can also apply them. According to one illustration embodiment of the system of this invention, and a method, picture reduction modification transforms all the pictures in a page by one of 1 set of predetermined scaling factors like 25%, 50%, and 75%, and makes the reduced picture the hypertext link which returns to an original image.

[0048] To picture reduction modification, in addition, all the abbreviation modification (Elide All transform), Three syntactic abbreviation modification of modification (First Image Only transform) and bookends modification (Bookends transform) was also developed for pictures only the first picture. In all the abbreviation modification, all the pictures are omitted from a document. By modification, all the pictures other than the first picture are omitted from a document only the first picture. In bookends modification, all the pictures other than the picture of the beginning and the last are omitted from a document. When the omitted picture has the available "ALT" (specified) text of HTML, it is transposed to each text. Or when an ALT text cannot be obtained, the omitted

picture is transposed to a standard icon. The ALT text or standard icon to the picture by which each ministry abbreviation was carried out is also made into the hypertext link linked to an original image.

[0049] According to one illustration embodiment of the document re authoring system of this invention, and a method, when screen space is restricted too much, or when a client apparatus cannot display a picture, a picture is removed from a document. However, the removed picture is used as an anchor of a hypertext link via the client side image map. When such a picture is removed, it should be recognized that the website expressed by the HTML document may be drawn impossible [ navigation ]. In order to prepare for this, in one illustration embodiment of the document re authoring system of this invention, and a method, a hypertext link is extracted from such a picture and the modification which formats them into the link anchor shown by a text list is used. The label to a text list is extracted from the URL (Uniform Resource Locator) portion of the link from the "ALT" tag, when the "ALT" tag of HTML of an image map exists. When this modification removes a picture, it saves the link attached to the picture for navigation.

[0050] It seems that the process of determining of which modification combination is applied to arbitrary pages, on the whole, needs the artistic capability of human being of a certain gestalt at first according to arbitrary client displays. However, the automatic document re authoring system and method of this invention gain much heuristics used by manual re-authoring training, and achieve success quite good for creation of the good page of the appearance doubled with arbitrary displays.

[0051] Individual page modification is set in order by each desirability. In order to determine of which modification combination should be applied to arbitrary documents, the document re authoring system and method of this invention perform depth first search of a document modification space using much heuristics which describes the precondition over the combination of modification and modification. Depth first search ensures that the "sufficiently good" version of a document is found by using the combination of most desirable modification. When more desirable modification is inapplicable, or only when fully not reducing a document, inferior modification is used in desirability.

[0052] The document re authoring system and method of this invention search a document modification space by the best priority method. Each state of this search space expresses the version of a document, and an initial state expresses a script "as authoring was carried out." The number showing the measure of the merit showing the quality of the document in the state is added to each state. The measure, i.e., the valuation function, or the value of the merit to each state is a rough estimate of a screen field required for the display of the whole document of the document in the state. One state can be developed to the following state by applying single modification art to the re-authoring finishing document in the state.

[0053] The most promising document state, i.e., the state where the present required (current) viewing area is the minimum, is chosen for every step of a search process, and modification is applied, in order to transform a document into the document state which is promising from the present state if possible. Shortly after the state where a "sufficiently good" document version is included is created, search can be suspended, and the document version is returned and drawn by the client apparatus. Or search is continued until all the contents of the original page are included or expressed to 1 set of sub pages good enough. Even if all searched, when a document version good enough is not found, the best document found during search is returned and drawn by the client apparatus. When restrictions of hard size are not satisfied with the best document, either, the more destructive modification which divides a document in the middle of a paragraph is applied.

[0054] Drawing 3 shows what different modification applied to the document 200 becomes the re-authoring finishing sub pages 210, 220, and 230 of a different result. Depending on the information given to the system and method of this

invention, one of the sub pages 210, 220, and 230 will be chosen by the user as a "best" re-authoring finishing page. . And said that the sub pages good enough for the contents removed, for example from the 1st sub pages were created. When re-authoring is required, or when there are still no best sub pages "well enough", and when, Re-authoring of the best re-authoring finishing sub pages 210, 220, or 230 which applied additional modification to the sub pages produced from the best selected re-authoring finishing sub pages 210, 220, or 230, or were chosen can be carried out further.

[0055]heuristics information including determination when the precondition over the turn and each modification art which various modification art is applied to a given state and a document version, or sub pages "are good enough" is used at several places by the document re authoring system and method of having followed this invention. Generally, the modification which carries out the minor change of the document is liked better than the modification changed more into large-scale. For example, the direction which reduces a picture 25% is more preferred than reducing 75%.

[0056]The precondition of each modification art specifies the modification and other modification in which combination is possible. For example, it is meaningless even if it applies both all the outline-izing and the 1st sentence abbreviation to the same document. A precondition also specifies the contents of the document in which art is going to be applied, and the necessary condition on structure. For example, all the outline-sized modification should be applied only when at least three section headers exist in the document under re-authoring. The present conditions for "being good enough" are quite simple. That is, search will be suspended if a field required for a document or sub pages becomes a predetermined multiple of the screen field of a client display. Generally, this predetermined multiple is larger than one, and it is 2.5 in one illustration embodiment. The multiple of the higher one of this assumes that a user does not mind that only a few scrolls a display to one way.

[0057]The contents of the document can be divided into two or more smaller "sub pages", if modification is applied to a document as shown in drawing 1 as a result. However, in order for each of these sub pages to download and display on a client, it may still be too large. In order to prepare for this problem, the document re authoring system and method of this invention create the list of the sub pages created by each sequence of modification attached to the state of expressing the document version of a result. Once the version of the document only called version of the first sub pages actually distributed to a client good enough good enough is chosen, The list of the created sub pages to the version is added to the global-area list which listed the page of the waiting for re-authoring. And the document re authoring system and method of this invention carry out re-authoring of each of these waiting pages for re-authoring until they can distribute all the sub pages of a result to a client. This procedure is shown in the following pseudo codes, and refer to the above-mentioned best priority re-authoring process over a single input page for "reauthor."

```
Digestor (initialpage).tobereauthored= {initialpage}.todeliver={} while
(tobereauthored!={}) nextpage=pop (tobereauthored) bestversionstate=reauthor
(nextpage).todeliver.append(bestversionstate.page)
tobereauthored.append(bestversionstate.subpages) return todeliver[0058]All the
re-authoring finishing sub pages are stored in cash as a modification parsing
tree. If a user navigates a changed document and demands sub pages, a
corresponding parsing tree will be drawn and it will be sent to a client.
```

[0059]When the document re authoring system and method of this invention carry out re-authoring of the document, they analyze a document first and constitute the parsing tree of a document, or abstract syntax tree (AST) expression. Next, the document re authoring system and method of this invention apply a series of modification to a parsing tree. And the document re authoring system and method of this invention map each modification parsing tree produced as a result, and return it to a document display, and the document display can become a different

document format from the format of the inputted script.

[0060] Document modification is realized using a standard procedure. The condition function which will return "truth" if a standard procedure incorporates the state node in a document version space and the modification should be applied to the state, and a new document version, In order to create the new state where the measure of new quality and the sub pages of a result are included, the function of operation called when actually applied to the state where the modification occurs is included. Modification of the following three types can be defined.

- 1) What is always performed on a page before a planning process start.
- 2) What is used for the best priority planning process.
- 3) What is always performed on a page before being changed so that it may return to a surface type type (surface form) like HTML from the last abstract syntax tree.

[0061] In the state where modification was applied, modification operates a parsing tree, in order to create the new version of a document. The operation The 5th international World-Wide-Web meeting, France, Paris, It is similar to a thing given in "interactive reconstruction (Interactively Restructuring HTML Documents) of an HTML document" of S. BONOMU and others in May, 1996 (S. Bonhomme et al.). If a parsing tree is omitted or transformed selectively, the HTML hypertext link which certainly refers to the node identifier of the subtree of all the influenced parsing trees will be added to a parsing tree, and a user enables it to require the original portion of the document changed during re-authoring.

[0062] The document re authoring system and method of this invention take record of the combination of the modification already tried on the assumption that all the modification was commutative via the global-area list of modification SETSU 65 TO, in order to ensure that the state where it overlapped is never constituted.

[0063] As mentioned above, one illustration the document re authoring system and the method of having followed this invention were realized as a HTTP proxy server. A HTTP proxy server receives the demand to an HTML document, a document being taken out from the specified HTTP server, an HTML document being analyzed and a parsing tree or an abstract syntax tree being constituted from a taken-out HTML document, and label attachment of each parsing tree node being carried out, and by a peculiar identifier, If required, arbitrary embedded images will be taken out so that the size of the taken-out picture can be judged. Once this is completed, the document re authoring system and method of this invention will be initialized in the state where the parsing tree to the taken-out script is included. The document re authoring system and method of this invention between each re-authoring cycle, The state of having the best document version in the place till then is chosen, the modification art in which the best application is next possible is chosen, the selected modification is applied, and, as a result, a new state and a new document version are created. It is made into a premise for the convolution of modification to be always commutative, and in order to ensure that the state where it overlapped is not constituted, some checks are used by the document re authoring system and method of this invention.

[0064] According to one illustration embodiment of the document re authoring system of this invention, and a method, 15 kinds of modification art were realized. Namely, all the outline-izing (FullOutline), outline-izing to H1 (OutlineToH1), Outline-izing (OutlineToH2) to H2, outline-izing to H3 (OutlineToH3), Outline-izing (OutlineToH4) to H4, outline-izing to H5 (OutlineToH5), Outline-izing (OutlineToH6) to H6, the 1st sentence abbreviation (FirstSentenceElision), 25% of picture reduction (ReduceImage25%) and 50% of a picture reduce (ReduceImage50%), 75% of picture reduction (ReduceImage75%), all the picture abbreviations (ElideAllImages), They are a bookends picture (BookendImages) which leaves only the first picture (FirstImageOnly), and font size reduction (ReduceFontSize).

[0065] This illustration embodiment of the re-authoring software system of this invention and the method was realized with the Java programming language. In addition to functioning as a true proxy server, this HTTP proxy server system can also answer the demand to specific URL which has the document drawn up by the HTTP proxy server itself. This is used in order to provide a user with the control based on the form of the HTTP proxy server, the document re authoring system, and the method. This illustration embodiment of a document re authoring system can be processed without even a very complicated page taking 2 seconds using the Java JIT compiler of Symantec (Symantec) on 200 MHz Pentium.

[0066] The user of the document re-authoring software system of this invention and a method having to do first is directing the font size of the default browser font which specifies and uses display size according to the device to be used. This information is required in order to estimate the required screen field of a text block. In order to succeed in this, the form 300 which a user demands specific control URL from a HTTP proxy server, and is shown in drawing 4 as a result is distributed.

[0067] Once a user constitutes a document re authoring system, he can start extraction of the document from a distributed network like World Wide Web. The original page 400 and the re-authoring finishing page 410 which are shown in drawing 5 are illustrating the re-authoring capability of the document re authoring system of this invention, and a method. In this example, the document re authoring system of this invention and this illustration embodiment of the method chose using 25% of picture reduction combining the 1st sentence abbreviation, and drew the page 410 displayed from the original page 400. And the re-authoring finishing page 410 is displayed on the browser window 420. In the re-authoring system of this invention, and this illustration embodiment of a method. The user can demand trace of a re-authoring session by requiring another control URL from a HTTP proxy server immediately after page extraction, in order to judge which modification was applied.

[0068] Drawing 6 shows one illustration embodiment of the environment 500 by which the automatic document re authoring system, the method and/or automatic document filtering system, and method of this invention are realized. As shown in drawing 6, the environment 500 contains the device 510 of the limited viewing area which has a display which has the viewing area dramatically restricted compared with the viewing area of the monitor for a desktop or laptop computers. As shown in drawing 6, the environment 500 contains further the transceiving equipment communications system 550, the host node 570 of a distributed network, and the remaining portion 590 of a distributed network.

[0069] Probably, in the environment 500, the device 510 of the limited viewing area is usually personal digital assistance (PDA), a cellular phone, etc. which were connected to the transceiving equipment communications system 550 by the radio channel 530. Therefore, as shown in drawing 6, the device 510 of the limited viewing area will usually contain the antenna 520, and, on the other hand, the transceiving equipment communications system 550 will usually contain the corresponding antenna 540. The device 510 of the limited viewing area will usually communicate with the transceiving equipment communications system 550 via the radio channel 530 using the radio frequency signal transmitted between the antenna 520 and 540.

[0070] The transceiving equipment communications system 550 changes into an usable form the analog or digital signal received from the device 510 of the limited viewing area via the communication channel 530 by the host node 570 of a distributed network. And the transceiving equipment communications system 550 outputs the signal received by the host node 570 of a distributed network via the communication link 560 via the communication channel 530. It should be recognized that an arbitrary communication [ in which the communication link 560 has the capability to transmit a suitable signal between the transceiving equipment communications system 550 and the host node 570 of a distributed network ] structure which are publicly known or were developed recently may be sufficient.

An exact structure of the transceiving equipment communications system 550 and the communication link 560, Since it becomes a problem of the design selection depending on how these elements are realized, it does not carry out describing these elements further, but for a person skilled in the art, such design selection becomes clear easily and anticipation attaches it.

[0071] It should also be recognized that the device 510 of the limited viewing area can connect with the host node 570 of a distributed network also by means other than radio channel 530 like the communication link 522. That is, other publicly known arbitrary communication structures, such as modem connection through a Local Area Network, a Wide Area Network, a public exchange telephone network, or a cable television system, may be sufficient as the communication link 522. For example, the user of the device 510 of the limited viewing area could connect the device 510 of the limited viewing area instead of communication through the radio channel 530 to the public exchange telephone network using the modem. Therefore, a user will call to the host node 570 of a distributed network directly.

[0072] It is not concerned with whether the host node 570 of a distributed network is connected to the device 510 of the viewing area restricted how eventually, When the demand to the document transmitted to the device 510 of a viewing area with which the host node 570 of the distributed network was once restricted is received, the host node 570 of a distributed network, First, it is judged whether the demanded document is located near the host node 570 of a distributed network. When the demanded document is not located in the neighborhood, the host node 570 of a distributed network communicates with the remaining portion 590 of a distributed network via the communication structure 580, in order to require a document. The specific node of the remaining portion 590 of the distributed network which stores the document receives the demand from the host node 570 via the communication structure 580 eventually, and returns the demanded document to the host node 570 via the communication structure 580. It should be recognized that an arbitrary communication [ by which the communication structure 580 links mutually the node located in the wide area of a distributed network ] structure which are publicly known or were developed recently, and a protocol system may be used.

[0073] If the host node 570 of a distributed network receives the once demanded document, The HTTP proxy server performed on the host node 570 of a distributed network carries out re-authoring of the demanded document based on the information about the device 510 of the limited viewing area provided beforehand. And the first re-authoring finishing page is transmitted to the device 510 of the limited viewing area by the host node 570 via either the wireless communications link 530 or the communication link 522. A user may judge that it is necessary to see the additional information removed from the re-authoring finishing page, if the distributed page is examined. In this case, a user sends a demand to the host node 570 of a distributed network via either the wireless communications link 530 or the communication link 522, in order for desired re-authoring finishing sub pages to come to hand. The host node 570 answers this demand and transmits to the device 510 of the viewing area to which the further re-authoring finishing sub pages of the script were restricted via either the radio channel 530 or the communication link 522.

[0074] Drawing 7 shows the flow of this information more to details. As shown in drawing 7, the user of the device 510 of the limited viewing area, A user sends the demand to a specific document to the HTTP proxy server 571 which exists on the host node 570 of a distributed network from the device 510 of the limited viewing area to examine the specific document which exists on a distributed network. Next, the HTTP proxy server 571 transmits to the specific remote node 591 on the distributed network which stores the page of which the demand to a specific document was required. The specific remote node 591 returns the demanded script to the document re authoring system 600 which exists on the HTTP proxy server 571. Re-authoring of the document re authoring system 600 is carried out to two or more sub documents which can display a script on the device 510 of a

viewing area with which it was restricted to each as exactly as possible. And the document re authoring system 600 is distributed to the device 510 of the viewing area to which the first re-authoring finishing page was restricted, and, on the other hand, other re-authoring finishing sub pages are stored in the re-authoring finishing sub-pages cash 636 of the document re authoring system 600. Therefore, when the user of the device 510 of the limited viewing area wants to see the information which exists on one of the re-authoring finishing sub pages stored in the re-authoring finishing sub-pages cash 636, A user makes the demand to the sub pages transmit to the device 510 of the limited viewing area. The sub pages stored in the demanded cash are distributed to the device 510 of the viewing area limited from the re-authoring finishing sub-pages cash 636.

[0075]Although HTTP server 571, the document re authoring system 600, and the re-authoring finishing sub-pages cash 636 are shown in drawing 7 as an independent element, Realizing as a portion [ like the module in which single application software differs generally ] in which these elements are and from which a simple substance differs should be recognized.

[0076]Drawing 8 is a functional block diagram showing more the outline of one illustration embodiment of the document re authoring system 600 in details. As shown in drawing 8, the document re authoring system 600, Each is mutually connected via data / control bus 680 including the controller 610, I/O interface 620, the memory 630, the abstract syntax tree generating circuit 640, the document size weighting network 650, the modification circuit 660, and the re-map circuit 670 from a tree to a document. As for the communication links 522, 560, and 580 mentioned above about drawing 6, each is connected to I/O interface 620.

[0077]The memory 630, . Include the original page memory part 631, the display size memory part 632, the abstract syntax tree memory part 633, the search space part 634, the modification memory part 635, the re-authoring finishing page cash 636 mentioned above about drawing 7, and the waiting sub-pages list 637 for re-authoring. A separate portion is included functionally [ many ]. The original page memory part 631 stores the returned script which was returned from the remote node 591 of the distributed network which stores the page demanded by the device 510 of the limited viewing area.

[0078]In order to obtain various parameters about the device 510 of the limited viewing area used by the document re authoring system 600 in order that the display size memory 632 may carry out re-authoring of the page according to the device 510 of the specific limited viewing area, Many form documents used by the document re authoring system 600 are stored. The specific size parameter to the device 510 of at least one limited viewing area also stores the display size memory 632. It should be recognized that there is a potential way a large number differ in realization of the document re authoring system 600, to various parameters about the device 510 of the limited viewing area. According to one illustration embodiment, the document re authoring system 600 can store various parameters to the device 510 of the specific limited viewing area, only while the device 510 of the limited viewing area is continued and connected to the document re authoring system 600. In this case, whenever re connection of the device 510 of the specific limited viewing area is carried out to the document re-authoring 600, the document re authoring system 600 will send various forms used in order to obtain various parameters about the device 510 of the limited viewing area, and, Whenever it accesses the document re authoring system 600 first, the user can ask for re-supply of these various parameters.

[0079]The above-mentioned thing reduces size required for the display size memory 632, While the arbitrary systems for recognizing the device 510 of the specific limited viewing area are not needed, This system needs the process of automating supply of the information from the device 510 of the restricted viewing area to which the big burden was applied by the user of the device 510 of the limited viewing area to the document re authoring system 600. This automation could be provided by, for example, requiring the information from the device 510 of a viewing area that the document re authoring system 600 was restricted. It is

inputted by the user during the session before information is already with the document re authoring system 600, Probably, the user does not need to be engaged in re-supply of the information on the re-authoring system 600 in operation, when stored on the device 510 of the viewing area to which the information was then restricted.

[0080]Or the information is storable in the display size memory 632 with the recognition code which can be made to supply from the device 510 of a viewing area with which it was restricted to the user at the time of a session start with the document re authoring system 600. It will be lost that it can ask for all the re-supplies of various parameters about the device 510 of the viewing area limited whenever the user accessed the document re authoring system 600 by supplying a recognition code to the document re authoring system 600 again.

[0081]Anyway, as mentioned above the document re authoring system 600, When carrying out re-authoring of the original page stored in the original page memory 631, it is made to be settled using various parameters about the device 510 of the limited viewing area as exactly as possible on the small viewing area of the device 510 of a viewing area with which each re-authoring finishing page was restricted.

[0082]The abstract syntax tree memory part 633 stores the abstract syntax tree generated by the abstract syntax tree generating circuit 640 from the script stored in the original page memory 631. The modification memory part 635 stores above-mentioned various modification, and stores similarly the conditions about the ability not to use [ / which and which ] it together between the conditions which can apply each modification, and modification of versatility. The modification memory 635 also stores the display of desirability when applying arbitrary specific modification to a specific original page or a re-authoring finishing page. That is, various modification has a general order which reduces a picture for more restrictive modification to which only a small quantity reduces a picture in large quantities, or is emphasized rather than more excessive modification which removes a picture thoroughly as mentioned above.

[0083]The re-authoring finishing page cash 636, A document size weighting network based on various parameters about the device 510 of the limited viewing area stored in the display size memory 632, If the abstract syntax tree corresponding to a specific re-authoring finishing page or sub pages shows a sufficiently good thing, the abstract syntax tree corresponding to a re-authoring finishing page or sub pages is stored. The waiting sub-pages list 637 for re-authoring stores the abstract syntax tree to the sub pages created by modification of the sub pages of a script or a front state. These sub pages will generally contain the picture of arbitrary size reduction images or arbitrary omission images, and all the texts of the arbitrary text segment which had the omitted contents.

[0084]The script by which the search space memory 634 is finally stored in the original page memory 631, Or when applying the various modification stored in the modification memory 635 based on the specific state of the search space under present operation to various sub pages stored in the waiting sub-pages list 637 for re-authoring, many states created by the modification circuit 660 are stored.

[0085]Especially each state i in the search space 634 contains an evaluation value part, a modification abstract syntax tree part, and a sub-pages list part. An evaluation value part stores the evaluation value generated to the re-authoring finishing page or sub pages corresponding to the state i where it was generated by the document size weighting network 650. A modification abstract syntax tree part stores the modification abstract syntax tree to the state i where it is created by the modification circuit 660, by applying one of the modification in the modification memory 635 to the parent of a state to the state i. When a sub-pages list part applies the specific modification used in order that the modification circuit 660 may create the state i, it stores the list of the sub pages created since the contents of the arbitrary origin removed from the page corresponding to the state i were stored.



[0086] It should be recognized that the state 0 corresponds to the script stored in the original page memory 631. Especially the evaluation value part of the state 0 corresponds to the evaluation value generated to the script which has not carried out re-authoring at all. In this state 0, the modification abstract syntax tree part stores the original abstract syntax tree which was generated by the abstract syntax tree generating circuit to the script and which does not change. Since the script includes all the original information before the state 0 and sub pages are finally unnecessary, a sub-pages list will be called empty.

[0087] Drawing 9 is illustrating various states where it is stored in the search space memory part 634. Especially drawing 9 shows the document which comprised one section header, one text paragraph, and one picture. As shown in drawing 9, in the state of [ 0 ] the initial state, the script has not changed yet. This initial state also shows the evaluation value generated to the original estimate, i.e., a script. Drawing 9 also shows the state 1 where it was created from the state 0, by applying "all the picture abbreviation" modification to the document of the state 0. As shown in the state 1, the re-authoring finishing sub pages of the state 1 contain a section header and a text, but a picture is not included. Rather, the re-authoring finishing sub pages of the state 1 include the link which is linked to the sub pages which store in the place of a picture the picture omitted from the re-authoring finishing sub pages of the state 1 in the re-authoring finishing page of the state 1 and which was displayed as "IMG." The state 1 also shows the evaluation value to this re-authoring finishing document. As shown in drawing 9, the required size of the re-authoring finishing page has dropped to 1/4 of the required size of the original page by which re-authoring is not carried out.

[0088] Drawing 9 also shows the two additional states 2 created by applying other modification to the document of the state 0, i.e., a state, and the state 3. Finally, drawing 9 shows the three additional states 4 created by applying additional modification to the sub pages of the re-authoring finishing document of the state 1, or the state 1, i.e., a state, the state 5, and the state 6. For example, although displayed [ sub pages including a picture ] on the device 510 of the limited viewing area, when too still large, In order to obtain a re-authoring finishing document good enough, the middle sub pages created by applying "25% of picture reduction", "50% of picture reduction", or "75% of picture reduction" modification to a picture will be displayed on the device 510 of the limited viewing area.

[0089] At this time, the document re authoring system 600 of drawing 8 receives the returned script via the communication link 580 in operation. The received script is inputted via I/O interface 620, and is stored in the original page memory 631 under control of the controller 610. Next, under control of the controller 610, the abstract syntax tree generating circuit 640 inputs the script from the original page memory part 631, and generates an abstract syntax tree from a script. And the abstract syntax tree generated by the abstract syntax tree generating circuit 640 is stored in the abstract syntax tree memory part 633 of the memory 630 under control of the controller 610.

[0090] Next, the document size weighting network 650 under control of the controller 610, Various parameters about the device 510 of the specific limited viewing area with which the re-authoring finishing document from the abstract syntax tree corresponding to the script stored in the original page memory 631 and the display size memory 632 is returned are inputted. And the document size weighting network 650 generates an evaluation value, and stores the evaluation value in the state 0 of the search space memory part 634. The document size weighting network 650 also outputs the display of whether it is good for outputting to the device 510 of a viewing area with which the document of the state 0 was restricted to the controller 610 via one side of the communication link 522 or 560 enough. When a script is good already enough, a script is returned immediately, without changing further.

[0091] Next, the modification circuit 660 inputs the document in the state 0 where

it is expressed under control of the controller 610 by the abstract syntax tree to the state 0, and applies one of the modification stored in the modification memory 635 to the abstract syntax tree in the state where it was inputted. Especially the modification circuit 660 determines whether the selected modification should be first applied to the present state i of a document for the present state i. For example, as mentioned above, when the present state i of a document does not include a picture at all, even if it applies picture reduction or abbreviation modification to the state of this document, nothing becomes. In order to acquire the present state i of a picture, when abbreviation" modification is already altogether applied except the picture of "beginning, even if it applies abbreviation" modification to this present state i altogether except the picture of "beginning and the last, nothing becomes.

[0092]The modification circuit 660 applies the modification to the abstract syntax tree to the state, and creates the child state j on the assumption that the present modification with the selected modification circuit 660 can apply to the present state i of a document expressed by the modification abstract syntax tree to the present state i properly. The child state j is provided with the following.

The abstract syntax tree which changed.

The sub-pages list in which sub pages required in order to reach this child state j with the waiting for modification are shown based on the contents omitted from the script.

Finally, the document size weighting network 650 determines whether to be good for outputting to the device 510 of a viewing area which evaluated the document obtained in the state of [ j ] the child and with which the document of the result was restricted under control of the controller 610 enough. And the evaluation value is stored in the newly created child state j.

[0093]After the modification circuit 660 creates the new child state j, in order to evaluate the required size of the document corresponding to the state j, the modification abstract syntax tree to the state j is outputted to the document size weighting network 630.

[0094]If the abstract syntax tree to the page of the beginning of the document which changed is determined that it is good enough, the abstract syntax tree will be outputted to the re-map circuit 670 from a tree to a document, and the circuit 670 will draw the first re-authoring finishing sub pages from the abstract syntax tree. The first re-authoring finishing sub pages are outputted to I/O interface 620 from the re-map circuit 670 from a tree to a document, and are transmitted to the device 510 of the eventually limited viewing area. Simultaneously, the modification circuit 660 continues application of additional modification to the arbitrary sub pages of the result which transformed the script into the first re-authoring finishing sub pages good enough. If such sub pages are transformed into sub pages separately good enough, the abstract syntax tree to such individual sub pages good enough, It is stored in the re-authoring finishing page cash 636 until the demand to the sub pages from the device 510 of the limited viewing area is received by the document re authoring system 600.

[0095]If the demand to the sub pages is received by the document re authoring system 600, The abstract syntax tree to the demanded sub pages is outputted to the re-map circuit 670 from a tree to a document, and the re-map circuit 670 draws the demanded re-authoring finishing sub pages from an abstract syntax tree. The demanded re-authoring finishing sub pages are outputted to I/O interface 620 from the re-map circuit 670 from a tree to a document, and are transmitted to the device 510 of the eventually limited viewing area.

[0096]It should be understood that each of the circuit shown in drawing 6-8 and other elements can be realized as a portion of the general purpose computer programmed appropriately. Or each circuit shown in drawing 6-8 physically in one or ASIC beyond it (dedicated integrated circuit) as separate hardware circuitry, Or it is realizable using a discrete logic element or a discrete circuit element, using FPGA, PDL, PLA, or PAL. The specific gestalt of each circuit shown in drawing 6-8 is included in design selection, for a person skilled in the art,

becomes clear easily and can be expected.

[0097]The links 522, 560, and 580 the device 510 of the limited viewing area to the host node 570. Or it is [ for connecting the host node 570 to the transceiving equipment communications system 550 or the remaining portion 590 of a distributed network ] publicly known, or it should be recognized that the arbitrary devices or system developed recently may be used. Therefore, the links 522, 560, and 580 are realizable as connection through direct cable connection, a Wide Area Network, or a Local Area Network, the connection through intranet, or connection through the Internet, respectively. Generally, the arbitrary connection systems or structure usable for connecting a corresponding device to the host node 570 via a distributed network which is publicly known or was developed recently may be sufficient as the links 522, 560, and 580.

[0098]It should be recognized further that the document re authoring system 600 is preferably realized on the programmed general purpose computer. However, the document re authoring system 600, The programmed microprocessor or microcontroller as a dedicated purpose computer and a peripheral-integrated-circuits element, And it is realizable also on programmable logical devices, such as the hard wiring electronic circuit or logic circuit like ASIC or other integrated circuits, a digital signal processor, and a discrete device circuit or PLD, PLA, FPGA, or PAL, etc. The arbitrary devices which have the realization capability of a finite state machine, that is, generally have the realization capability of the flow chart shown in drawing 11-14 can be used for realization of the document re authoring system 600.

[0099]The memory 630 shown in drawing 8 is preferred, and is static, or is realized using a dynamic RAM. However, the memory 630 is realizable even if it uses a floppy disk and a disk drive, the optical disc that can be written in and a disk drive, a hard drive, a flash memory, other arbitrary volatility that are publicly known or were developed recently, or non-volatile alterable memory. One which stores the control program for the controller 610, or the portion beyond it can be further included by the memory 630. Generally such a control program preferably, It is stored using other variable [ arbitrary ] or eternal nonvolatile memory which are publicly known or were developed recently using nonvolatile memory, such as a flash memory, ROM, PROM, and EPROM or EEPROM, using CD-ROM and a disk drive.

[0100]Drawing 10 shows the abstract syntax tree generated from another illustration script and its document. As shown in drawing 10, a document contains one picture, one table which has two lines x three rows, and one text paragraph. The abstract syntax tree of the result generated from this page contains the root node by which label attachment was carried out with "Page." Three intermediate nodes, "Image", "Table", and "Paragraph", which correspond to each of a picture, a table, and a text paragraph individually are prolonged from the "Page" node of the route. As shown in drawing 10, two intermediate nodes which correspond to each of two lines individually, "Row 1" and "Row 2", are prolonged from the middle "Table" node. At the last, three nodes respectively and individually corresponding to three cells of each line are prolonged from each of "Row 1" and "Row 2" node.

[0101]The modification applied first will be replacing a picture as large as life with the node which expresses a reduction image 25% generally, for example, in order to carry out re-authoring of the page shown in drawing 10. And the new abstract syntax tree which has a root node corresponding to a full-scale picture will be formed, and it will be linked to the node of the picture to which the modification abstract syntax tree was reduced by the hypertext link. The picture reduction modification removed thoroughly will be applied [ picture / 50% reduction, 75% reduction, and ] to the script one by one in the picture until it acquires a picture good enough, when the re-authoring finishing page which has a reduction image 25% is not good still enough. In each case, an abstract syntax tree will include the link to another abstract syntax tree which includes a full-scale picture from the modification node corresponding to a picture. Even if it removes a picture thoroughly, when still insufficient for becoming a

re-authoring finishing document good enough, With the application of front modification, a table can be transformed into 1 set of each linked cells as mentioned above, or a text paragraph can be moved to separate sub pages with the application of the 1st sentence abbreviation modification.

[0102]Drawing 11 and 12 are flow charts which show the outline of the one illustration method for page re-authoring according to this invention. As shown in drawing 11 and 12, control is started at Step S100, it continues to Step S110, and a user connects to the re-authoring system according to this invention the device which has the limited viewing area. Next, in Step S120 a re-authoring system, In order to acquire the required information about the device of the limited viewing area so that re-authoring of the page demanded according to the display of the device of the limited viewing area can be carried out, one or the parameter form beyond it is transmitted to a user. And in Step S130, a re-authoring system inputs the parameter information from a user, and stores the inputted parameter information in a memory. And control continues to Step S140.

[0103]As mentioned above about drawing 6 and 7, the parameter information collection process which showed the outline at Steps S120 and S130 is automatable so that a user may not be engaged in execution of Steps S120 and S130 in operation. Or as shown in the option step S135, Steps S120 and S130 can be replaced by Step S135. In Step S135, the device of the viewing area which the user inputted into the re-authoring system in operation, or was restricted outputs automatically the identification code which identifies the parameter information stored before receiving the device of this specific limited viewing area. And control continues to Step S140 also here.

[0104]In Step S140, the demand to the document on a distributed network is outputted to a re-authoring system from the user who is using the device of the limited viewing area. And in Step S150, a re-authoring system obtains the document demanded from the distributed network. Next, in Step S160, in order to build the abstract syntax tree of the document which came to hand, a document is analyzed. And in Step S170, the evaluation value to the document which came to hand is generated from an abstract syntax tree. And control continues to Step S180.

[0105]In Step S180, it is determined whether to be good for displaying on the device of a viewing area with which the evaluation value was analyzed, and the document which came to hand did not carry out re-authoring at all, but \*\* was also restricted enough. When good enough, control is jumped to Step S340. When that is not right, control continues to Step S190.

[0106]In Step S190, one or the reserve re-authoring modification beyond it is applied to the abstract syntax tree of the script which came to hand. This reserve re-authoring modification is used, for example in order [ of a script ] to remove only no portions which consume a viewing area excluding the contents. For example, such a portion of the document which came to hand includes the banner and other graphical elements which show the link of a part of another page or page. The picture without these contents is replaced by the text link. However, since such modification removes no contents from a picture actually, preservation of the portion removed in such page re-authoring is unnecessary. The portion which can be removed without affecting the contents of the script contains in others space and the format command which adds the esthetic format without other contents to a script. Another modification which changes various fonts of a document into a single standard font, and finally reduces the unnecessary viewing-area demands of a complicated big font is applicable.

[0107]If reserve re-authoring modification is applied at Step S190, control will continue to Step S200 and the evaluation value to a reserve re-authoring finishing script will be generated. Next, in Step S210, the evaluation value of a reserve re-authoring finishing document is checked, and it is determined whether to be good for displaying on the device of a viewing area with which the reserve re-authoring finishing document was restricted enough. When good enough, control

is jumped to Step S340 also here. When that is not right, control continues to Step S220.

[0108] In Step S220, the state 0 of the search space corresponding to a reserve re-authoring finishing document is chosen as the present state of a search space. Next, in Step S230, the first modification is chosen as the present modification. And in Step S240, it is determined whether the present modification can apply to the abstract syntax tree of the present state. As the outline was described above, whether each of modification of versatility being able to apply effectively in the re-authoring finishing document of the present [ modification / the ] and the present modification have the conditions which show [ the modification applied before that, and ] whether combination is properly possible. When the present re-authoring finishing document corresponding to the present state says that the present modification is effectively contradictory to the application possibility of and no modification applied before that, control continues to Step S250. When that is not right, control is jumped to Step S290.

[0109] In Step S250, the present state is transformed into a child state using the present modification, and the child state of the result containing a modification abstract syntax tree and the sub pages of arbitrary results is added to a search space. And in Step S260, the evaluation value to the document corresponding to the modification abstract syntax tree corresponding to the child state created at Step S250 is generated. Next, in Step S270, an evaluation value is analyzed and it is determined whether to be good for the document corresponding to the child state created at Step S250 displaying on the device of the limited viewing area enough. When an evaluation value shows that a re-authoring finishing document or sub pages is sufficiently good, control is jumped to Step S310. When that is not right, control continues to Step S280.

[0110] In Step S280, it is judged whether all the modification was applied to the present state. When all modification has not been applied yet, control continues to Step S290. When that is not right, control is jumped to Step S300.

[0111] In Step S290, the next modification is chosen as the present modification, and control is jumped to Step S240 and returns. On the other hand, in Step S300, the state of a search space of having the best evaluation value is chosen as the present state. And control is jumped to Step S230 and returns.

[0112] In Step S310, the document or sub pages specified by the present state is added to re-authoring finishing page cash as the first re-authoring finishing page suitable for distribution to the device of the limited viewing area which has advanced the demand, or the next re-authoring finishing page. And in Step S320, it is judged whether arbitrary sub pages arose from the sub pages good enough added to re-authoring finishing page cash. When there are such sub pages that still need re-authoring, control continues to Step S330. When that is not right, control is jumped to Step S340.

[0113] In Step S330, the state of the search space corresponding to one of the waiting sub pages for re-authoring is chosen as the present state. And control is jumped to Step S230 and returns. When there are on the other hand already no sub pages which need re-authoring, it is Step S340 and the first re-authoring finishing page is outputted to the device of the limited viewing area which has advanced the demand. And a control routine is completed at Step S350.

[0114] Drawing 13 shows the outline of one illustration embodiment of abbreviation modification according to this invention. As shown in drawing 13, an abbreviation modification routine is started at Step S400, it continues to Step S410, and the portion by which the present page or sub pages are removed is chosen. And in Step S420, it is copied to sub pages with a new selected portion. Next, the identifier to the selected portion is created in Step S430. Generally, an identifier is created, using the contents of the selected portion a little. For example, when the selected portion is a paragraph or other text strings, an identifier serves as a portion of the beginning of the 1st sentence of the selected text part, or the 1st sentence. When the selected portion is a picture, the identifier can

become some texts used for identifying a picture in a web document. And control continues to Step S440.

[0115] In Step S440, the sub pages which had a present page or sub pages created, and the link to link are created. And in Step S450, the selected portion is removed from a present page or sub pages, and an identifier and a link are added to the present page. A control routine stops at the following step S460.

[0116] Drawing 14 shows the outline of one illustration embodiment of front modification according to this invention. As shown in drawing 14, front modification is started at Step S500, it continues to Step S505, and the table of the highest level is chosen as the present table. Next, in Step S510, the present table is checked and it is judged whether the table made into nests arbitrary in the present table exists. If it exists, control will continue to Step S515. When that is not right, control is jumped to Step S520. In Step S515, one of the tables made into the nest is chosen as the present table as the new present table. And control is jumped to Step S510, and returns, and it is judged whether the table made into the nest exists in the table made into the nest selected as this present table.

[0117] If the table made into the nest does not exist in the present table, the present table is checked at Step S520, and it is judged whether a sidebar exists in the present table. In existing, control continues to Step S525. When that is not right, control is jumped to Step S535. In Step S525, the list of links is created from all the links in all the sidebars of the present table. Next, in Step S530, the list of links is arranged at the last of the present table. And control continues to Step S535.

[0118] The present table is divided into two or the portion beyond it in Step S535. As especially mentioned above, one method of dividing the present table into a portion is dividing each cell of a table into an individual portion. And in Step S540, each portion of the present table is copied to separate new sub pages, and the "next" and a "pre-" link are added to such each sub pages. Next, the present table is replaced with 1 set created at Step S540 of linked sub pages in Step S545. And control continues to Step S550.

[0119] In Step S550, the present table is checked and it is judged whether it is a table of the highest level. When it is not a table of the highest level, the table of an upper level exists from that [ at least one ] which still needs to be divided into a portion. Therefore, control continues to Step S555. When that is not right, control is jumped to Step S560.

[0120] In Step S555, a table including the present table is chosen as the new present table. And control is jumped to Step S510, and returns, and it is judged whether the table made into the nest still exists in the present table. On the other hand, a control routine is completed at Step S560.

[0121] Drawing 15 is a flow chart which shows the outline of one illustration embodiment of picture reduction modification according to this invention. starting picture reduction modification at Step S600 -- Step S610 -- then, the picture in the present sub pages reduced is chosen. Next, a reduction image is created based on the reduced coefficient relevant to the specific picture reduction modification under application. Next, in Step S630, it is judged whether the present sub pages were analyzed and the selected picture was reduced before. If reduced in front, control will be jumped to Step S670. When that is not right, control continues to Step S640.

[0122] In Step S640, the selected picture is copied to new sub pages. Next, the link to new sub pages is created in Step S650. And in Step S660, in order to remove a full-scale picture from a present page or sub pages and to form a re-authoring finishing page, a reduction image and the created link are added to the present page. And control is jumped to Step S680.

[0123] On the other hand, in Step S670, a full-scale picture is not moved from the present sub pages, but the old picture reduced before is removed from the present sub pages, and a new reduction image is added to the present sub pages. However, since the present sub pages should already have a link to sub pages including the

full-scale picture created before, they do not need to create the new sub pages which add a link to the present sub pages again, and store the full-scale picture. And control continues to Step S680 and a control routine ends it there.

[0124] Since the very small only for [ a text ] type display is used for the cellular phone, By being clear using perfect automatic re-authoring of a document, when there is only too much information in a typical web document, it does not become the leisure when web browsing of a cellular phone is pleasant or useful rash. Typically, especially these devices and services will be used in order to find and display the information which the user is looking for. That is, these devices and services will be used for the search and extraction of information which set up the target. A user enables it to extract only the interesting portion of a document via the easy end user script language to which the document filtering system and method of this invention combine a structural page navigation command with regular representation pattern matching and a report writing function.

[0125] The 7th international World-Wide-Web meeting, Australia, Brisbane, R. mirrors (R.) in April, 1998 "sphinx of Miller et al.: The framework for personal site specific web crawler creation (site-specific SPHINX: a framework for creating personal) The sphinx system of a statement provides Web crawlers" with the visual tool with which a user can create a web crawler of-special-make "personal" functionally similar to the system of this invention, and the filtering mechanism of a method. The 7th international World-Wide-Web meeting, Australia, Brisbane, A. SUGIURA et al. in April, 1998 (A.) "Internet scrapbook of Sugiura et al.: Automation (Internet Scrapbook: automating Web browsing.) of the web browsing work by demonstration programming The Internet scrapbook of a statement to tasks by programming-by-demonstration", If a user chooses the element from a web page visually and a web page is changed, the function similar to the page element search to the specific page of the system of this invention and a method which can update those elements in a "scrapbook" is provided. Some goods also provide the report generation of a company, or other applications like database parent population with similar functionality, for example. The headliner pro (Headliner Pro) of Lana Qom given in Lana Qom incorporation (Lanacom, Inc.) homepage <http://www.headliner.com>, And the center stage (CenterStage) of the one display of a statement to one display incorporation (OnDisplay, Inc.) homepage <http://www.ondisplay.com> both, The visual editor to which a user is made to direct which structural moiety of a web page is extracted is provided. However, a user is not provided with the capability for these systems to both extract the contents based on regular representation or a keyword.

[0126] The document filtering system and method of this invention have the capability for a user to extract partial information from a document based on the command written by the high level script language. The document filtering system and method of this invention, In addition to re-authoring of the extracted information using the above-mentioned document re authoring system and method of this invention, page structure navigation, regular representation collation, a site scan (web crawling), i.e., web crawling, and repeated type collation are combined.

[0127] A filter script is simply inputted into a text file, and is saved on a Web server. This filter script will be performed always, if a user demands that URL. The filter script loaded the target web page typically, and were structurally described by regular representation. The extracted contents are sent so that the specific location in the web page is scanned, and the contents found on those locations are extracted, and it may be properly formatted via a document re authoring system before being returned to a user and.

[0128] The document filtering system and method of this invention, By providing 1 set of easy HTML document navigation options using the concept of "the present context (current context)" in an HTML document, parsing tree generation and navigation of the document re authoring system of this invention and a method are used. The present context is similar to the "cursor" of database programming, and

refer to the location in an HTML document for it.

[0129]Actually, refer to the node in a HTML parsing tree for the present context. If it carries out the role which makes it move about this reference within a parsing tree and a desired portion is found in it until the portion of a request of an HTML document is found in a navigation command, it can extract a desired portion. For example, drawing 10 shows the HTML document and the parsing tree corresponding to it. When a document is first loaded by execution of the "GO URL" command, the present context has pointed out the root node of the parsing tree which refers to the whole document intrinsically.

[0130]Drawing 16 shows one illustration embodiment of the document re authoring system 600 which includes further the filter circuit 690 which realizes the document filtering system and method of describing an outline to this specification. Especially the filter circuit 690 inputs the demanded filter which was demanded by the user via one side of the communication link 522 or 560 under control of the controller 610. A filter is supplied via the communication link 580 from the node of the distributed network which stores such a filter. Next, the filter circuit 690 filters the demanded document, in order to input the document demanded from the node of the distributed network which stores the demanded document and to extract the demanded page element. The filter circuit 690 is stored in the place where the script was stored in this extracted page element at the beginning [ of the original page memory 631 ]. And the document re authoring system 600 operates this extracted page element as if it was a script by which re-authoring is carried out.

[0131]When extracting a page element from a script, the abstract syntax tree which was generated by the abstract syntax tree generating circuit from the script, and was stored in the abstract syntax tree memory 633 is used for the filter circuit 690.

[0132]Drawing 17 shows the outline of one illustration embodiment of the flow of information in the case also of also being filtered in the demanded document. As shown in drawing 17, after the demand to a filter is outputted to the HTTP proxy server 571 by the device 510 of the limited viewing area, the demand to a filter, It is transmitted to the remote node 592 of the new distributed network which stores the demanded filter by the HTTP proxy server 571. The remote node 592 which stores the demanded filter returns the demanded filter to the document filter 690. Next, the document filter 690 requires the document from the remote node 591 of the distributed network which stores the demanded page under control of the controller 610. The remote node 591 which stores the demanded page returns the document to the document filter 690. And the document filter 690 filters the returned document using the filter returned from the remote node 592, and the abstract syntax tree generated by the abstract syntax tree generating circuit 640. The document filter 690 returns the extracted page element to the document re authoring system 600, and the page element extracted by the system 600 is treated like the script by which re-authoring is carried out as mentioned above.

[0133]What goes into the present context in order to choose as a page navigation command the contents specified more (into), What comes out of the present context, surrounds (out) (enclosing), and tends toward structure, And there are three types of what scans a page sequentially from the beginning of the present context for navigating to a certain kind of following structure (next) which may be properly included in the present context, for example, or may not be included.

[0134]The simplest type of a navigation command goes into the present context. For example, in the document and the present context which are shown in drawing 10, execution of a command "GO ROW 2" will move the present context to the object of the 2nd line of the table in the present context, as shown in drawing 18.

[0135]Expansion of the present context is also possible. That is, a parsing tree can also be gone up toward a root node by using "GO ENCLOSING." For example, in the document shown in drawing 18, and a context, the result of the "GO ENCLOSING



TABLE" command serves as the present context shown in drawing 19.

[0136] Finally, the present context can move between the objects in a page so that it may be visible to a user one by one for forward or backward at movement. This is attained before or by moving backward within the prefix (prefix) scan of a parsing tree from the present location in the present context. As a result, search is first performed within the present context and it is continued by the object which follows the present context on a page next. For example, the "GO PREVIOUS IMAGE" command moves to the picture before being found one by one from the present context.

[0137] In addition to the page element which had named, a navigation command can also be specified using regular representation. For example, a "GO NEXT "DOW\*\*sJONES\*\*s\*(\*\*d+) \*\*s\*POINTS"" command moves to collation next to the regular representation of specification of the present context using the prefix scan of the text block on a page. The filtering system and method of this invention can separate a subexpression, and can call them into an output sequence.

[0138] An easy above-mentioned navigation command can also be used for navigating between 1 set of linked web pages by using a "LINKEDPAGE" page object type. For example, the "GO FIRST LINKEDPAGE" command moves to the hypertext link of the beginning in the present context, loads the page referred to, and moves the present context to the route of the parsing tree of the document. On the other hand, the "GO ENCLOSEDLINKEDPAGE" command returns the present context to the hypertext link led to the document under present processing.

[0139] The scan between pages is coped with by the stack of script starting to which each script starting makes script state information (the present context is included) specific URL and a parsing tree, and a pair. This provides the quick navigation to between [ the linked pages ] order, and it is called for so that the "GO ENCLOSED LINKEDPAGE" command may be supported.

[0140] If the present context is moved to the page object of the object of interest, the "REPORT" command will be used in order to extract it. The "REPORT" command can be emitted several times within one filter script, and the page element extracted in that case is connected. The "REPORT" command can be used also for insertion of the arbitrary sequences to an output, and can include the sub sequence from regular representation pattern matching. For example, a "REPORT"Dow:\*\*1"" command adds the sub sequence identified by the sequence "Dow:" and identifier "1" which were extracted by the output of the filter during regular representation collation.

[0141] A user does not sometimes often know how many a specific kind of page element exists on a web page. For example, as for the number of paragraphs of the news story of the online magazine (e-zine) of daily publication, not understanding beforehand is common. Lack of this information is equipped with the "FOREACH" command by performing the sequence of a command for every page element with which it is satisfied of the specified standard which was found within the present context. If this is used with the "LINKEDPAGE" target, the functionality of web Spider (web spider) which can visit all the linked pages in one website is provided. The apostrophe expresses the sequence of the effective filter command with the following examples.

[0142] The "FOREACH PARAGRAPH" command moves to each paragraph in the context of the present of DO--END one by one, and executes the specified command.

[0143] The "FOREACH LINKEDPAGE" command loads each page which can reach one by one via a hypertext link from the present page of DO--END, and executes the specified command.

[0144] A filter contains failure in navigation, failure in regular representation collation, or a web page retrieval error. When arbitrary kinds of error is encountered, a filter is started always repeatedly next to the "FOREACH" loop of the No. 1 inside where the command acting as an obstacle (offending) is embedded simply. When an error arises in the highest level of a filter, a filter

suspends execution and creates arbitrary halfway outputs.

[0145]The document re-authoring software system and method of this invention achieve success to document automatic re-authoring for displaying on the device which has a small screen. One illustration embodiment of the document re authoring system of this invention and a method has been privately tested on the wide range page corresponding to many screen sizes. The document re authoring system of this invention and this illustration embodiment of the method created the output which can be navigated [ that decipherment is possible and ].

[0146]According to one illustration embodiment, the document re authoring system and method of this invention total a space required for all the pictures and texts simply, in order to obtain the estimate of a screen field required for a document. Although this is suitable for the quite high-density document of a minimum configuration like the document of a Xerox annual report, unfilled space does not commit it well in plenty in a certain document, for example, the document using advanced layout art as shown in a table. According to a 2nd illustration embodiment, when the document re authoring system and method of this invention format each document version on a viewing area, they contain the size S tee meter (estimator) which performs many of work performed by the browser. Elements other than a required screen field, such as the actual necessary width of a re-authoring document and necessary bandwidth to a user not liking horizontal scrolling, and esthetic measure, may also need to be included.

[0147]The user adjusts various heuristics used by the document re authoring system and method of this invention according to liking of him. For example, a user specifies relative liking of modification art, or a certain modification hopes that it can be specified that it does not use at all. liking [ as opposed to / abstract a high level more and / 1 set of trade-offs in a user ] of one -- " -- more -- the contents -- many -- " -- a pair -- " -- I hope that a display can be expressed more like [ it is large and ] ". I hope that it moves to a client side and the document filtering system and method of this invention can be made together with a browser so that various modification can be applied and canceled dynamically, until it obtains the result which a user satisfies.

[0148]The automatic document re authoring system of this invention, a method, and especially the illustration embodiment of an above-mentioned HTTP proxy server are preferably realized on the programmed general purpose computer. However, the automatic document re authoring system, the method, and especially the above-mentioned HTTP proxy server of this invention, A dedicated purpose computer, the programmed microprocessor or a microcontroller, and a peripheral-integrated-circuits element, It is realizable also on programmable logical devices, such as the hard wiring electronic circuit or logic circuit like ASIC or other integrated circuits, a digital signal processor, and a discrete device circuit, PLD, PLA, FPGA, or PAL, etc. Generally, arbitrary devices with the realization capability of a finite state machine can be used for realization of the automatic document re authoring system of this invention, a method, and an especially above-mentioned HTTP proxy server.

[0149]The automatic document re authoring system and method of having followed this invention, It can perform via plug-in to the conventional web browsers, such as calling the stand-alone re-authoring programs executed on an above-mentioned HTTP proxy server, or Netscape Navigator.

[0150]Although the automatic document re authoring system and method of this invention have been described about re-authoring of the document which came to hand from World Wide Web, The automatic re-authoring system and method of this invention, It can use also for re-authoring of the document which came to hand from arbitrary distributed networks, such as a Local Area Network, a Wide Area Network, intranet or other arbitrary distributed processings, and a memory network.

[0151]Although this invention has been described in relation to the specific embodiment which described outline above, for a person skilled in the art, it is

clear that many substitution, correction, and change are clear. Therefore, the desirable embodiment of this invention indicated above has intention of an explanatory thing, and is not restrictive. Various change does without deviating from the main point and range of this invention.

---

#### DESCRIPTION OF DRAWINGS

---

##### [Brief Description of the Drawings]

[Drawing 1] It is a figure showing re-authoring of a document to one section list page and many section pages according to one illustration embodiment of the document re authoring system of this invention, and a method.

[Drawing 2] It is a figure according to one illustration embodiment of the document re authoring system of this invention, and a method showing the layout table in which re-authoring is possible in two or more linked cells.

[Drawing 3] It is a figure showing what it may change re-authoring into the re-authoring state where documents differ based on application of different modification, according to the re-authoring system of this invention, and one illustration embodiment of a method.

[Drawing 4] It is a figure showing one illustration embodiment of control form for supplying display information to the HTTP proxy server according to the document re authoring system and method of this invention.

[Drawing 5] It is a figure showing one illustration embodiment of re-authoring of an illustration document according to the document re authoring system and method of this invention.

[Drawing 6] It is a block diagram showing the outline of one illustration embodiment of this invention using the document re authoring system and method of this invention.

[Drawing 7] It is a block diagram showing the outline of one illustration embodiment of the flow of the document re authoring system of this invention, and the document of a method.

[Drawing 8] It is a functional block diagram showing the document re authoring system of this invention, and the outline of one illustration embodiment of a method.

[Drawing 9] It is a figure showing one illustration embodiment of the document version search space of the document re authoring system of this invention, and a method.

[Drawing 10] It is a figure showing the picture generated from one picture according to this invention, and one illustration embodiment of an abstract syntax tree.

[Drawing 11] It is a flow chart showing the outline of one illustration embodiment of the method for document re-authoring according to this invention.

[Drawing 12] It is a flow chart showing the outline of one illustration embodiment of the method for document re-authoring according to this invention.

[Drawing 13] It is a flow chart showing one illustration embodiment of a method which performs abbreviation modification according to this invention.

[Drawing 14] It is a flow chart showing one illustration embodiment of a method which performs front modification according to this invention.

[Drawing 15] It is a flow chart showing one illustration embodiment of a method which performs picture reduction modification according to this invention.

[Drawing 16] It is a functional block diagram showing the outline of one illustration embodiment of the document re authoring system 600 of this invention including document filtering according to this invention.

[Drawing 17] It is a figure showing one illustration embodiment of the flow of the document under document filtering according to this invention, and re-authoring.

[Drawing 18] In order to navigate the inside of the abstract syntax tree generated from the picture shown in drawing 10, it is a figure showing one illustration embodiment which used the document filtering system and method of this invention.

[Drawing 19] It is a figure showing the further navigation in the abstract syntax tree of drawing 10 according to the document filtering system and method of this invention.

[Description of Notations]

510 A device of the limited viewing area

571 HTTP proxy server

600 Document re authoring system

610 Controller

620 I/O interface

630 Memory

640 Abstract syntax tree generating circuit

650 Document size weighting network

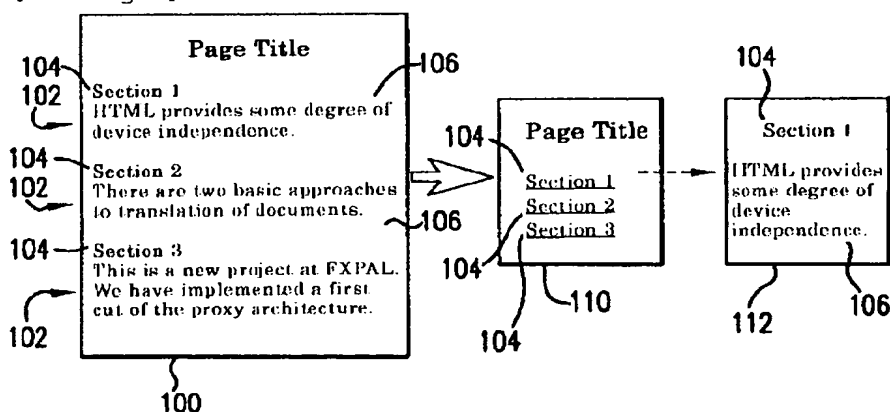
660 Modification circuit

670 The re-map circuit from a tree to a document

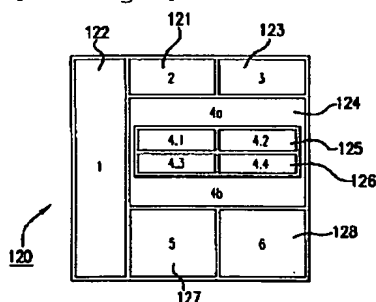
690 Document filtering subsystem

## DRAWINGS

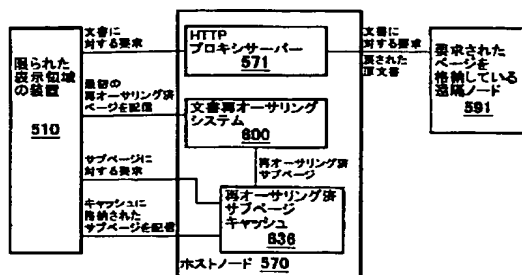
[Drawing 1]



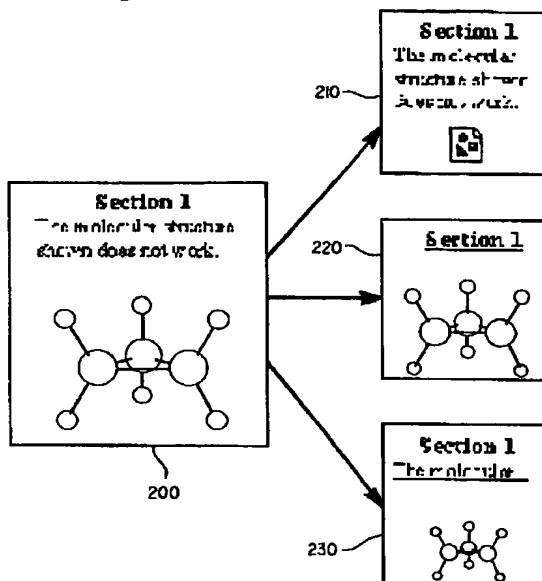
[Drawing 2]



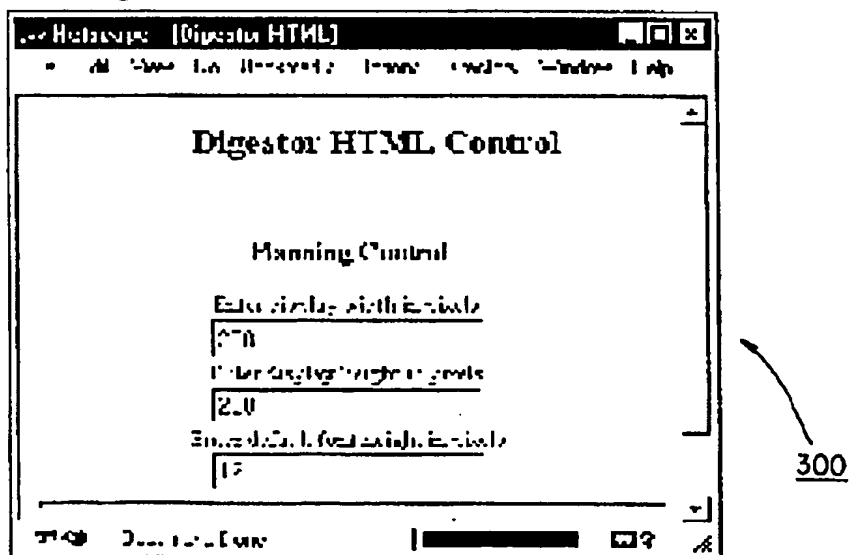
[Drawing 7]



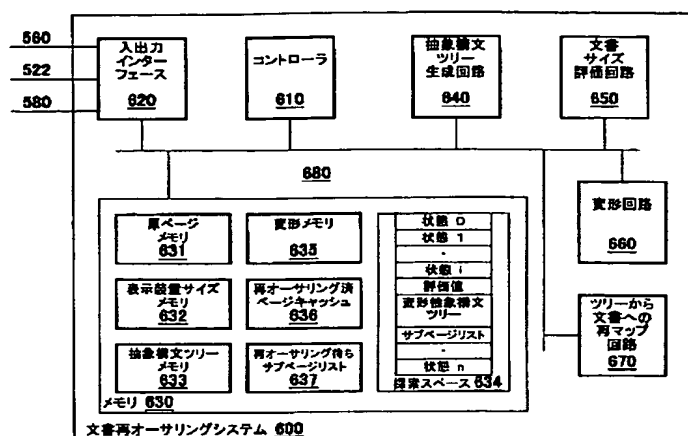
[Drawing 3]



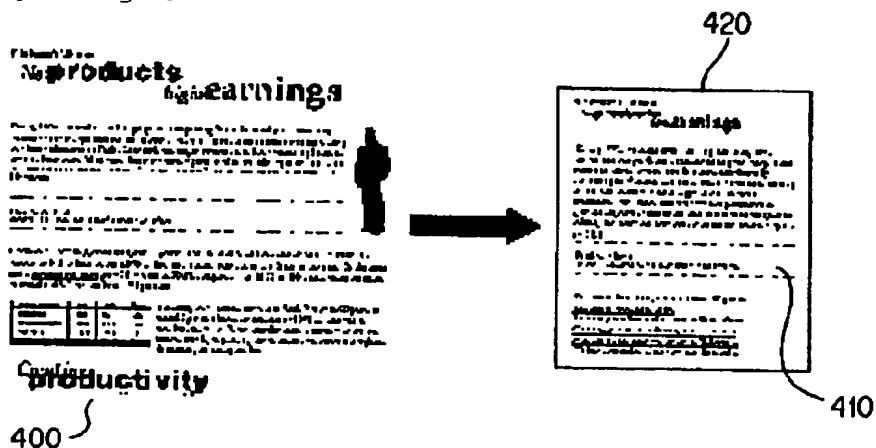
[Drawing 4]



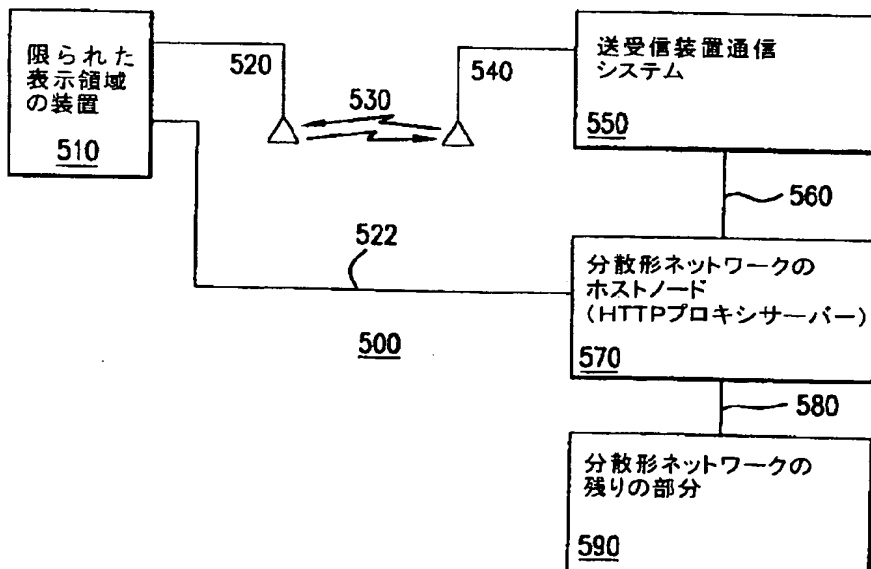
[Drawing 8]



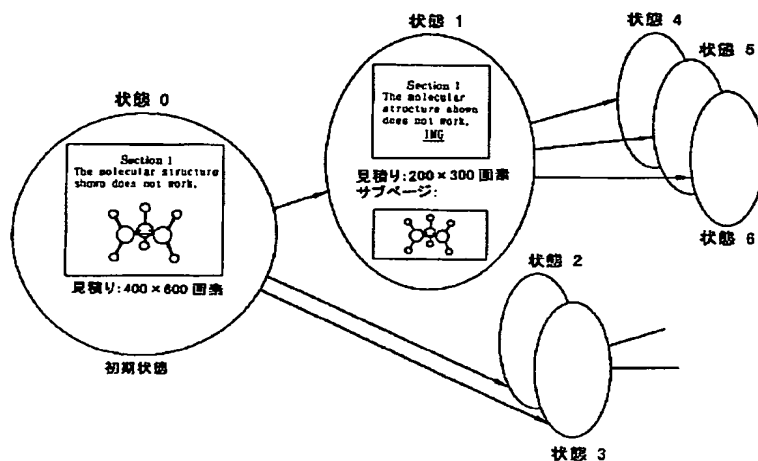
[Drawing 5]



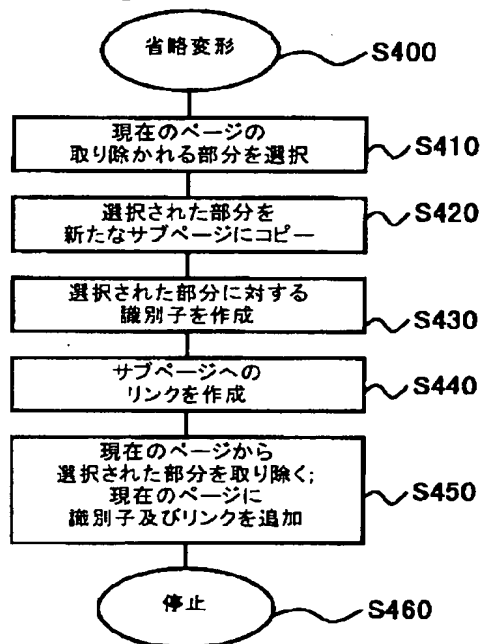
[Drawing 6]



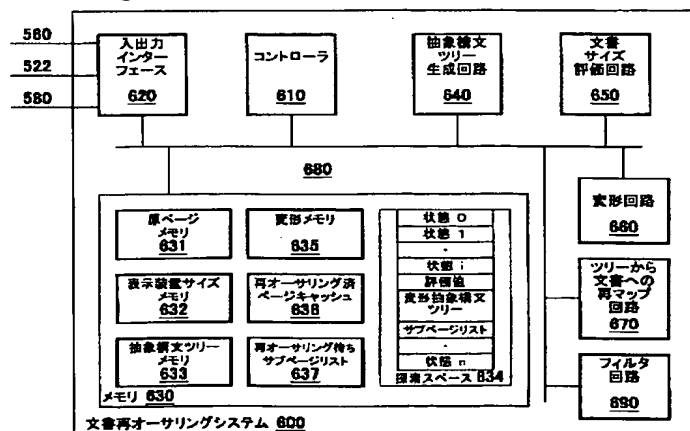
[Drawing 9]



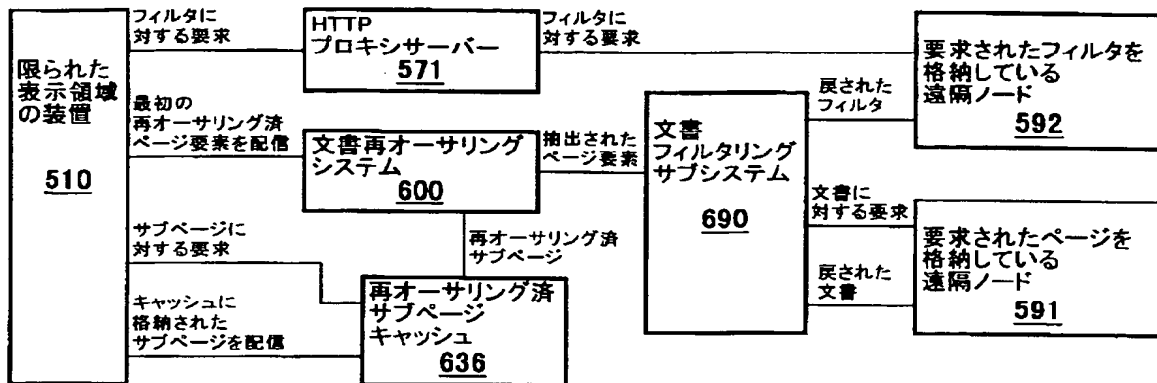
[Drawing 13]



[Drawing 16]

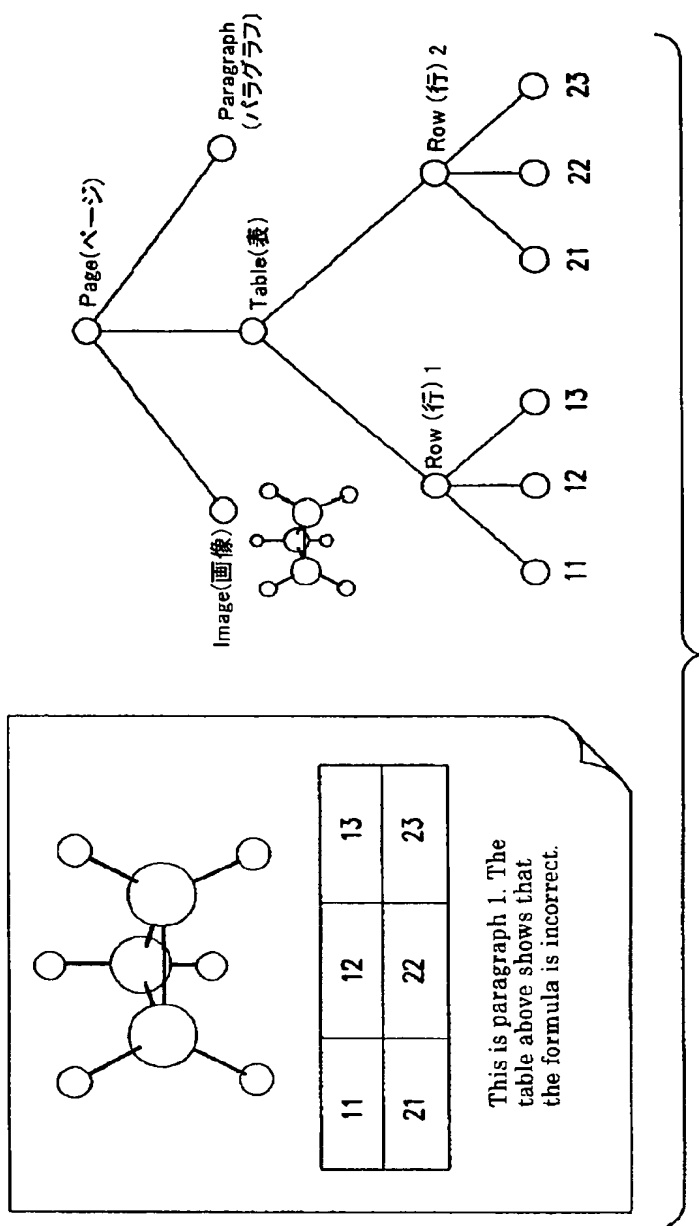


[Drawing 17]

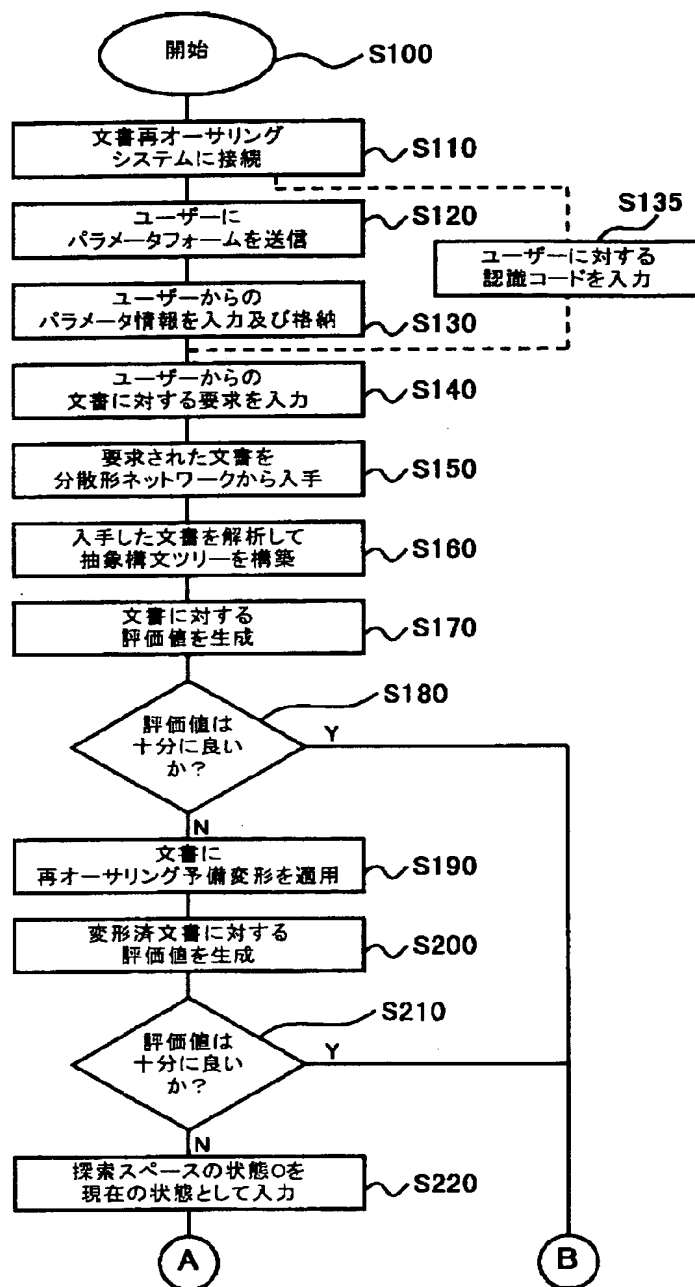


[Drawing 10]

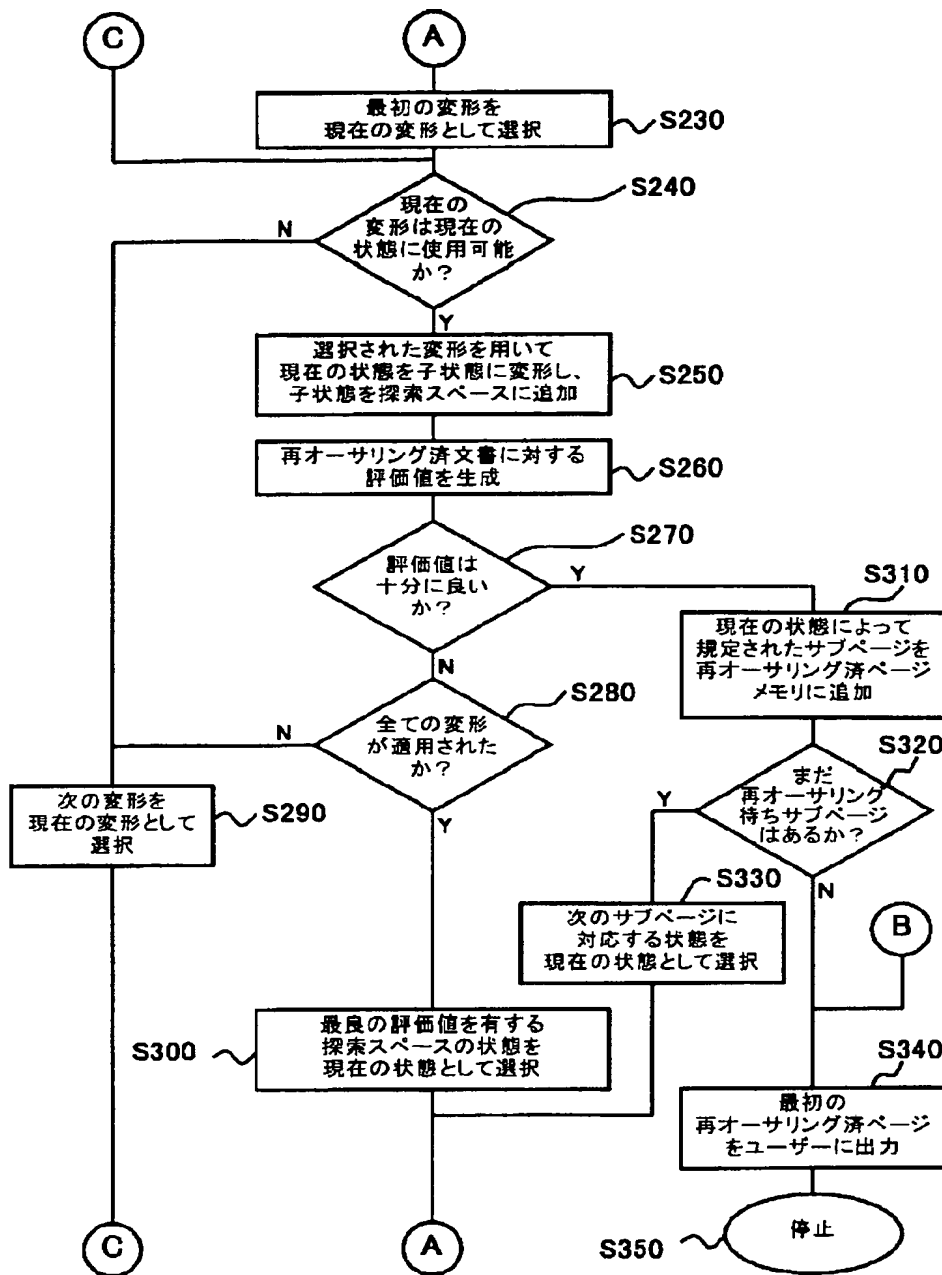




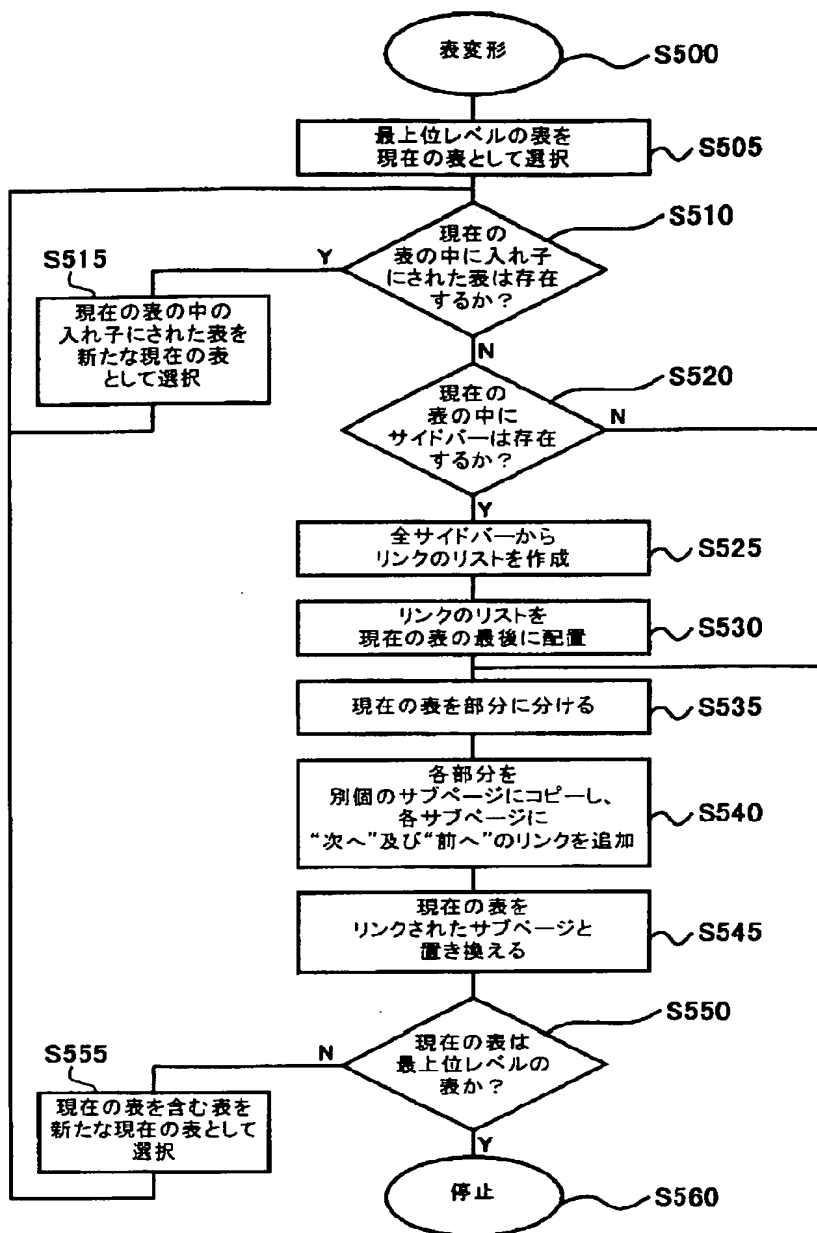
[Drawing 11]



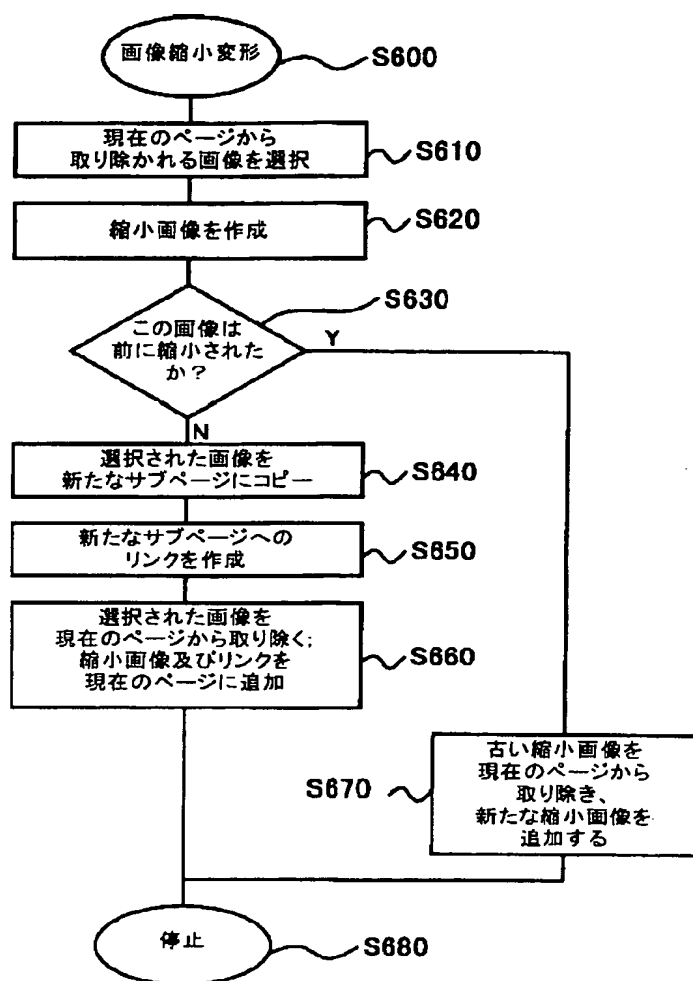
[Drawing 12]



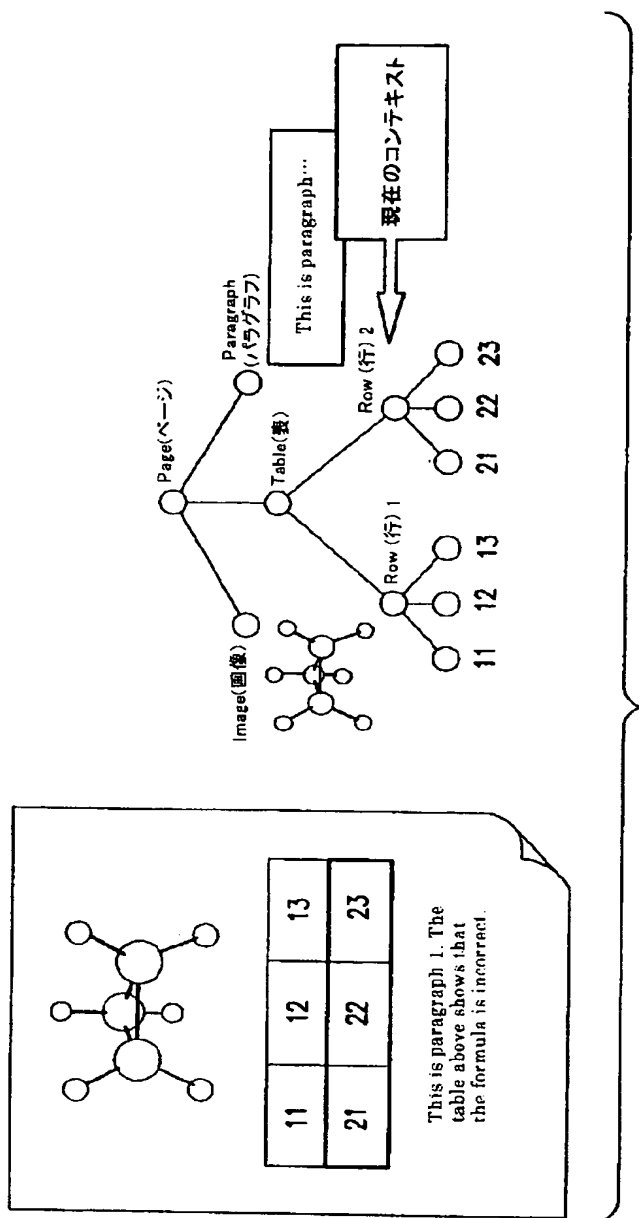
[Drawing 14]



[Drawing 15]



[Drawing 18]



[Drawing 19]

